

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2005, 17:57:15 ; Search time 176 Seconds  
(without alignments)  
2155.970 Million cell updates/sec

Title: US-09-684-890A-2

Perfect score: 3887

Sequence: 1 MAAAGPVGDELWQWLFPN.....GEHIREMVKQINDIRHNVNF 741

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt 03.\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	3817	98.2	741	1	NU88 HUMAN
2	3461.5	89.1	742	1	NU88 RAT
3	3452	88.8	753	1	NU88 MOUSE
4	2551.5	65.6	549	2	Q8BQF0
5	2417	62.2	728	2	Q707N0
6	2397	61.7	726	2	Q6DDV7
7	2386	61.4	726	2	Q707M9
8	2133.5	54.9	462	2	Q8CAD9
9	918	23.6	190	2	Q6Q7J4
10	701	18.0	633	2	Q7PX24
11	646.5	16.6	702	1	NU88 DROME
12	559.5	14.4	130	2	Q29177
13	318	8.2	342	2	Q7PKI7
14	248.5	6.4	810	2	Q9FFK6
15	211	5.4	890	2	Q7XPFO
16	184.5	4.7	803	2	Q9P382
17	175.5	4.5	888	2	Q7RXL7
18	166	4.3	980	2	Q9POK7
19	166	4.3	983	2	Q7Z514
20	163	4.2	787	2	Q8NAB1
21	163	4.2	980	2	Q7Z733
22	163	4.2	989	2	Q9P212
23	161	4.1	1253	2	Q86S00
24	158	4.1	344	2	Q9Y3T5
25	157	4.0	972	2	Q6VIA9
26	156	4.0	442	2	Q6AY97
27	155	4.0	609	2	Q8TXA4
28	154	4.0	964	2	Q8IIQ7
29	153	3.9	1366	2	Q6C134
30	152	3.9	920	2	Q7FAY1
31	149.5	3.8	442	2	Q9D8L5

32	149.5	3.8	894	2	Q7FAD5
33	149.5	3.8	2954	2	Q42263
34	147.5	3.8	978	1	RA50 AQUAE
35	146	3.8	786	2	Q6BGH4
36	145.5	3.7	1206	2	Q9NTC1
37	144.5	3.7	1249	2	Q8KIN2
38	143.5	3.7	793	2	Q8BKV3
39	143.5	3.7	898	2	Q9Z220
40	142.5	3.7	2663	1	CENE HUMAN
41	141.5	3.6	446	2	Q66H66
42	141.5	3.6	1302	2	Q80Y16
43	140.5	3.6	713	1	NU82 YEAST
44	140	3.6	876	1	RA50_METKA
45	140	3.6	1107	2	Q7RME7

#### ALIGNMENTS

#### RESULT 1

NU88 HUMAN STANDARD; PRT; 741 AA.  
 AC Q99567; Q9BWE5;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 23-OCT-2004 (Rel. 45, Last annotation update)  
 DE Nuclear pore complex protein Nup88 (Nucleoporin Nup88) (88 kDa nuclear pore complex protein).  
 GN Name=NUP88;  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND INTERACTION WITH NUP214/CAN.  
 RC TISSUE=Placenta;  
 RX MEDLINE=22388257; PubMed=9049309; DOI=10.1093/emboj/16.4.807;  
 RA Fornerod M., van Deursen J.M., van Baal S., Reynolds A., Davis D., Murti K.G., Franssen J., Grosveld G.;  
 RT "The human homologue of yeast CRM1 is in a dynamic subcomplex with CAN/Nup214 and the novel nuclear pore component Nup88.";  
 RL EMBO J. 16:807-816 (1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Lung;  
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heide F., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Brownstein M.J., Udén T.B., Tohiyuki S., Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J., Bosak S.A., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahey J., Helton E., Kettaman M., Madan A., Rodrigues S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalls D.E., Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences.";  
 RL Proc Natl. Acad. Sci. U.S.A. 99:16999-16903 (2002).  
 CC -!- FUNCTION: Essential component of nuclear pore complex.  
 CC -!- SUBUNIT: Interacts with NUP214/CAN.  
 CC -!- SUBCELLULAR LOCATION: Nuclear pore complex.  
 CC -!- TISSUE SPECIFICITY: Ubiquitous.

-----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
DR EMBL: Y08612; CAA69904.1; -;  
DR EMBL: BC000335; AAH00335.1; -;  
DR Genew; HGNC: 8067; NUP88.  
DR MIM; 602552; -;  
DR GO; GO:0005643; C:nuclear pore; TAS.  
DR GO; GO:0005215; P:transporter activity; TAS.  
KW Coiled coil; Nuclear protein; Protein transport; Transport.  
FT DOMAIN 585 651 Coiled coil (potential).  
FT CONFLICT 247 247 A -> D (in Ref. 1).  
FT CONFLICT 291 301 KLLGLGELPMHP -> KWAGSIAHAS (in Ref. 1).  
FT CONFLICT 456 456 K -> R (in Ref. 1).  
FT CONFLICT 518 518 P -> S (in Ref. 1).  
SQ SEQUENCE 741 AA; 83541 MW; 954A8E2E203BC20B CRC64;

Query Match 98.2%; Score 3817; DB 1; Length 741;  
Best Local Similarity 98.4%; Pred. No. 1.3e-205;  
Matches 729; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

QY 1 MAAAGPVGDGELWQTLNPNHVVFLRLREGLNQSPTEAEKPASSSLPSSPPQLLNRV 60  
DB 1 MAAAGPVGDGELWQTLNPNHVVFLRLREGLNQSPTEAEKPASSSLPSSPPQLLNRV 60  
QY 61 VFLGGLFLWDGESSFLVRLRGSGGEEEPALSOYQLLCINPPLFEIYOVLLSPQ 120  
DB 61 VFLGGLFLWDGESSFLVRLRGSGGEEEPALSOYQLLCINPPLFEIYOVLLSPQ 120  
QY 121 HHVALIGIGLWLELPKRWGNSEFEGGKSTVNCSTTPVAERFSTSLTKHAAWP 180  
DB 121 HHVALIGIGLWLELPKRWGNSEFEGGKSTVNCSTTPVAERFSTSLTKHAAWP 180  
QY 181 SEILDPHVLLTSDNVIIRIYSLREPOTPTNVIILSEAEESLVNKGRTAYTASLGATA 240  
DB 181 SEILDPHVLLTSDNVIIRIYSLREPOTPTNVIILSEAEESLVNKGRTAYTASLGATA 240  
QY 241 FFGPLDAPKTLFGQNGKDEVAVPLYLYENGETFLYILSHSPGNIMKAVGSIHA 300  
DB 241 FFGPLDAPKTLFGQNGKDEVAVPLYLYENGETFLYILSHSPGNIMKAVGSIHA 300  
QY 301 SAAEDNNGYDACAVALCLPCVPNLIIVATSGMLYHCVVLEGEEDDHTSEKSWDSRIDLI 360  
DB 301 SAAEDNNGYDACAVALCLPCVPNLIIVATSGMLYHCVVLEGEEDDHTSEKSWDSRIDLI 360  
QY 361 PSLYVFECEVELELAKLASGEDDPFSDPSCPVKLHRDPKCPSRHYCTHEAGVHSLGTW 420  
DB 361 PSLYVFECEVELELAKLASGEDDPFSDPSCPVKLHRDPKCPSRHYCTHEAGVHSLGTW 420  
QY 421 IHKLHFLGSDDEBKDSLOELSTEQKCFVEHILCTPLCPROPAPIRGFWIVPDILGPTM 480  
DB 421 IHKLHFLGSDDEBKDSLOELSTEQKCFVEHILCTPLCPROPAPIRGFWIVPDILGPTM 480  
QY 481 ICITSTYECILWPLSTVHPASPPLCTREDVEAESRLVLAETPDSFEKHIRSTLQPS 540  
DB 481 ICITSTYECILWPLSTVHPASPPLCTREDVEAESRLVLAETPDSFEKHIRSTLQPS 540  
QY 541 VANPAFLKASEKDIAPEPPECQLLSRATQVFREQYILKQDLAKBEIQRRVKLLCDQKKK 600  
DB 541 VANPAFLKASEKDIAPEPPECQLLSRATQVFREQYILKQDLAKBEIQRRVKLLCDQKKK 600  
QY 601 QLEDLSYCREERKSLREMAERLADKYEAKBQKEDIMNMKLLHSHFSELVPLSDSERD 660  
DB 601 QLEDLSYCREERKSLREMAERLADKYEAKBQKEDIMNMKLLHSHFSELVPLSDSERD 660  
QY 661 MKKELQLIPDQRLHGLGNAIKQVTMKDYQOQKMEKVLSPKPTIILSAYORKCIIQSILKE 720  
DB 661 MKKELQLIPDQRLHGLGNAIKQVTMKDYQOQKMEKVLSPKPTIILSAYORKCIIQSILKE 720  
QY 721 EGEHIREMVKQINDIRNHVNF 741

DB 721 EGEHIREMVKQINDIRNHVNF 741

RESULT 2  
NUP88\_RAT STANDARD; PRT; 742 AA.  
ID NU88\_RAT  
AC O08658;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 25-OCT-2004 (Rel. 45, Last sequence update)  
DE Nuclear pore complex protein Nup88 (Nucleoporin Nup88) (88 kDa nuclear pore complex protein) (Nucleoporin Nup84).  
GN Name=Nup88; Synonyms=Nup84;  
OS Rattus norvegicus (Rat);  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
SQ SEQUENCE FROM N.A., AND INTERACTION WITH NUP214/CAN.

RC TISSUE=Thyroid carcinoma;  
RX MEDLINE=97311070; PubMed=9166401; DOI=10.1083/jcb.137.5.989;  
RA Bastos R., Ribas de Pouplana L., Enarson M., Bodoor K., Burke B.;  
RT "Nup84, a novel nucleoporin that is associated with CAN/Nup214 on the cytoplasmic face of the nuclear pore complex.";  
RL J. Cell Biol. 137:989-1000(1997).  
CC -!- FUNCTION: Essential component of nuclear pore complex.  
CC -!- SUBUNIT: Interacts with NUP214/CAN.  
CC -!- SUBCELLULAR LOCATION: Nuclear pore complex.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; U93692; AAB52419.1; -;  
KW Coiled coil; Nuclear protein; Protein transport; Transport.  
SQ DOMAIN 597 652  
SQ SEQUENCE 742 AA; 83585 MW; E2C38D0DCBFFB3D8 CRC64;

Query Match 89.1%; Score 3461.5; DB 1; Length 742;  
Best Local Similarity 88.2%; Pred. No. 1.1e-185;  
Matches 655; Conservative 37; Mismatches 48; Indels 3; Gaps 2;

QY 1 MAAAGPVGDGELWQTLNPNHVVFLRLREGLNQSPTEAEKPASSSLPSSPPQLLNRV 58  
DB 1 MAAAGPVGDGELWQTLNPNHVVFLRLREGLNQSPTEAEKPASSSLPSSPPQLLNRV 60  
QY 59 NUVFGLGGLFLWDGESSFLVRLRGSGGEEEPALSOYQLLCINPPLFEIYOVLLSP 118  
DB 61 NLVFGGLGFLWDAGSAPLVRLRGSGGSEVPEPLSQYQRLCINPPLFEIHQVLLSP 120  
QY 119 TQHHVALIGIGLWLELPKRWGNSEFEGGKSTVNCSTTPVAERFSTSLTKHAAW 178  
DB 121 TQHHVALIGIGLWLELPKRWGNSEFEGGKSTVNCSTTPVAERFSTSLTKHAAW 180  
QY 179 YPSEILDPHVLLTSDNVIIRIYSLREPOTPTNVIILSEAEESLVNKGRTAYTASLGATA 238  
DB 181 YPSEILDPHVLLTSDNVIIRIYSLREPOTPTNVIILSEAEESLVNKGRTAYTASLGATA 240  
QY 239 VAFDGPGLDAPKTLFGQNGKDEVAVPLYLYENGETFLYILSHSPGNIMKAVGSIHA 298  
DB 241 VAFDGPGLDAPKTLFGQNGKDEVAVPLYLYENGETFLYILSHSPGNIMKAVGSIHA 299  
QY 299 HASAEDNNGYDACAVALCLPCVPNLIIVATSGMLYHCVVLEGEEDDHTSEKSWDSRID 358  
DB 300 MEPAEDNNGYDACAVALCLPCVPNLIIVATSGMLYHCVVLEGEEDDHTSEKSWDSRID 359  
QY 359 LIPSLYVFECEVELELAKLASGEDDPFSDPSCPVKLHRDPKCPSRHYCTHEAGVHSLGTW 418

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2005, 17:59:25 ; Search time 48 Seconds  
(without alignments)  
1485.346 Million cell updates/sec

Title: US-09-684-890A-2  
Perfect score: 3887  
Sequence: 1 MAABGPVGDGLWQTWLEN.....GEHIREWVKQINDIRNHVNP 741  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues  
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 79:.\*  
1: pir1:.\*  
2: pir2:.\*  
3: pir3:.\*  
4: pir4:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	158	4.1	345	2 T08700	hypothetical prote
2	149.5	3.8	2954	2 T14156	kinesin-related pr
3	147.5	3.8	978	2 A70387	conserved hypotet
4	142.5	3.7	2663	1 S28261	centromere protei
5	140.5	3.6	713	2 S56833	nuclear pore prote
6	140	3.6	1132	2 T00259	hypothetical prote
7	139.5	3.6	942	2 JC7316	testicular zinc fi
8	138.5	3.6	791	2 H72552	hypothetical prote
9	138	3.6	1169	2 A64505	pl15 homolog - Met
10	138	3.6	1875	2 S38173	myosin-like protel
11	137.5	3.5	1780	2 T17272	hypothetical prote
12	137	3.5	852	2 D72230	conserved hypotet
13	137	3.5	1955	2 T30934	myosin-like protei
14	135.5	3.5	886	2 H69378	conserved hypotet
15	135	3.5	1388	2 T30335	KLP2 protein - Afr
16	134	3.4	775	2 T37837	probable signal tr
17	134	3.4	1690	2 T13030	microtubule bindin
18	134	3.4	1957	2 A59294	skeletal myosin -
19	133.5	3.4	1217	2 A45493	phospholipase C-be
20	133	3.4	968	2 T51933	kinesin motor prot
21	132.5	3.4	1128	2 A86266	hypothetical prote
22	132	3.4	2116	2 A26655	myosin heavy chain
23	131.5	3.4	829	2 T01362	probable myosin he
24	131.5	3.4	853	2 T51505	hypothetical prote
25	131.5	3.4	1968	1 J05697	myosin heavy chain
26	131	3.4	631	2 JC4298	hyaluronan recepto
27	130.5	3.4	881	1 B71316	endopeptidase la (
28	130.5	3.4	1432	2 B85431	trichohyalin like
29	130.5	3.4	4725	1 A44357	dynein heavy chain

30	130	3.3	1030	2 T13163	Rab6 GTPase activa
31	130	3.3	1231	2 T18532	serine/threonine pr
32	129.5	3.3	757	2 B72273	hypothetical prote
33	129.5	3.3	1156	2 B70356	chromosome assembl
34	129.5	3.3	1999	1 S21801	myosin heavy chain
35	129.5	3.3	3084	1 MMSA	laminin alpha-1 ch
36	129	3.3	1312	1 BMBYDL	RAD50 protein - ye
37	129	3.3	1313	2 A48467	myosin heavy chain
38	129	3.3	4687	1 A39638	plectin - rat
39	127.5	3.3	756	2 T00367	hypothetical prote
40	127.5	3.3	952	2 T50451	hypothetical colle
41	127.5	3.3	1938	2 JC5421	smooth muscle myos
42	127.5	3.3	1972	2 JC5420	smooth muscle myos
43	127	3.3	558	2 D70449	conserved hypotet
44	126.5	3.3	943	2 T33795	hypothetical prote
45	126.5	3.3	1974	2 T30010	hypothetical prote

ALIGNMENTS

RESULT 1  
T08700  
hypothetical protein DKFZp564G013.1 - human (fragment)  
C;Species: Homo sapiens (man)  
C;Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 09-Jul-2004  
C;Accession: T08700  
R;Wambutt, R.; Heubner, D.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.  
submitted to the Protein Sequence Database, March 1999  
A;Reference number: Z16471  
A;Accession: T08700  
A;Molecule type: mRNA  
A;Residues: 1-345 <WAM>  
A;Cross-references: UNIPROT:Q9V3T5; EMBL:AL050011  
A;Experimental source: fetal brain; clone DKFZp564G013  
C;Genetics:  
A;Note: DKFZp564G013.1

Query Match 4.1%; Score 158; DB 2; Length 345;  
Best Local Similarity 24.6%; Pred. No. 0.021;  
Matches 58; Conservative 55; Mismatches 83; Indels 40; Gaps 8;

Qy	510	EDVEVAESSLRVLAE-----TPDSFEKHRSILQSRVANPAF-LKASEKDIAPPEE	560
Db	126	KEVEVAKLEKQLLEKAAAMDAMVPRSSYEKLSSESVLSLASKLKESVYKEKEKHSE	185
Qy	561	CLQLLSRATQVFEQYILKQDLAKEETQRRVLLCDQKKQLEDLSYCREERKSLREMAE	620
Db	186	VQIRSEVSQVKRE-----KENIQTLKSKQEVNELLQRFQQAQBELAMKRYAE	236
Qy	621	RLADKYEEAKEKQEDIMN-----RMKKLLHSFHSSELPVLSDSERDMKKELQLIPDQLRHLG	676
Db	237	S-SSKLEEDKDKKINEMSKVTKLKEALNSL-SQLSVYSTSSSKRSQSQLEALQQVQKQLQ	294
Qy	677	NAIKQVTMKDYQQQKMEKVLSPKPTIILSAVQRKCIQISILKEEGHIREMKVQI	732
Db	295	NQLAE--CKKQHOE-----VISVYRMHLLYAVOGQMDVDQVKLQKI	334

RESULT 2  
T14156  
kinesin-related protein - African clawed frog  
C;Species: Xenopus laevis (African clawed frog)  
C;Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 09-Jul-2004  
C;Accession: T14156  
R;Wood, K.W.; Sakowicz, R.; Goldstein, L.S.; Cleveland, D.W.  
Cell 91, 357-366, 1997  
A;Title: CENP-E is a plus end-directed kinetochore motor required for metaphase chromos  
A;Reference number: Z17893; MUID:98028574; PMID:9363944  
A;Accession: T14156  
A;Status: preliminary; translated from GB/EMBL/DBD  
A;Molecule type: mRNA  
A;Residues: 1-2954 <WOO>



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2005, 18:16:00 ; Search time 162 Seconds  
(without alignments)  
1775.601 Million cell updates/sec

Title: US-09-684-890A-2

Perfect score: 3887

Sequence: 1 MAAAGPVGDELWQTLN.....GHIHREVMVQINDIRNVNF 741

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1736639 seqs, 388188149 residues

Total number of hits satisfying chosen parameters: 1736639

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*

1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10D\_PUBCOMB.pep.\*  
17: /cgn2\_6/ptodata/2/pubpaa/US10E\_PUBCOMB.pep.\*  
18: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
19: /cgn2\_6/ptodata/2/pubpaa/US11A\_PUBCOMB.pep.\*  
20: /cgn2\_6/ptodata/2/pubpaa/US11\_NEW\_PUB.pep.\*  
21: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
22: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	285	7.3	62	9	US-09-864-761-40109
2	201	5.2	40	9	US-09-864-761-39809
3	194	5.0	858	16	US-10-437-963-140355
4	170	4.4	610	15	US-10-424-598-279883
5	166	4.3	980	15	US-10-211-462-145
6	163	4.2	787	15	US-10-104-047-3340
7	155	4.0	370	15	US-10-276-774-2414
8	152.5	3.9	849	15	US-10-424-599-279884
9	149.5	3.8	894	16	US-10-437-963-194984
10	147.5	3.8	978	17	US-10-732-923-3309
11	142.5	3.7	1196	15	US-10-282-122A-52737
					Sequence 40109, A
					Sequence 39809, A
					Sequence 140355,
					Sequence 279883,
					Sequence 145, App
					Sequence 3340, Ap
					Sequence 2414, Ap
					Sequence 279884,
					Sequence 194984,
					Sequence 3309, Ap
					Sequence 52737, A

#### RESULT 1

US-09-864-761-40109  
; Sequence 40109 Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Aecomica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30

#### ALIGNMENTS

12	142.5	3.7	2503	17	US-10-828-985A-11	Sequence 11, Appl
13	142.5	3.7	2543	17	US-10-828-985A-9	Sequence 9, Appl
14	142.5	3.7	2568	17	US-10-828-985A-7	Sequence 7, Appl
15	142.5	3.7	2663	16	US-10-723-860-749	Sequence 749, App
16	140	3.6	1130	15	US-10-260-708-67	Sequence 67, Appl
17	140	3.6	2099	16	US-10-723-860-267	Sequence 267, App
18	140	3.6	2390	15	US-10-092-900A-292	Sequence 292, App
19	139.5	3.6	1253	15	US-10-334-143-2	Sequence 2, Appl
20	138.5	3.6	696	16	US-10-408-765A-1067	Sequence 1067, Ap
21	138	3.6	1169	15	US-10-369-493-1095	Sequence 1095, Ap
22	138	3.6	1875	15	US-10-369-493-22285	Sequence 22285, A
23	138	3.6	1875	15	US-10-732-923-3334	Sequence 3334, Ap
24	138	3.6	1875	17	US-10-732-923-3335	Sequence 3335, Ap
25	137.5	3.5	798	14	US-10-032-585-7126	Sequence 7126, Ap
26	137.5	3.5	882	17	US-10-732-923-3300	Sequence 3300, Ap
27	137.5	3.5	1080	15	US-10-369-493-12560	Sequence 12560, A
28	137	3.5	852	17	US-10-732-923-3313	Sequence 3313, Ap
29	137	3.5	1955	17	US-10-732-923-3349	Sequence 3349, Ap
30	136.5	3.5	630	15	US-10-205-647A-6	Sequence 6, Appl
31	136.5	3.5	631	10	US-09-978-309A-48	Sequence 48, Appl
32	136.5	3.5	631	17	US-10-892-831-48	Sequence 48, Appl
33	136.5	3.5	745	15	US-10-108-260A-3547	Sequence 3547, Ap
34	135.5	3.5	623	16	US-10-416-330-38	Sequence 38, Appl
35	135.5	3.5	798	17	US-10-741-849-7169	Sequence 7169, Ap
36	135.5	3.5	886	15	US-10-369-493-1016	Sequence 1016, Ap
37	135.5	3.5	886	17	US-10-732-923-3310	Sequence 3310, Ap
38	135.5	3.5	1178	15	US-10-282-122A-52434	Sequence 52434, A
39	134	3.4	1979	14	US-10-205-823-419	Sequence 419, App
40	134	3.4	3113	16	US-10-723-860-862	Sequence 862, App
41	134	3.4	3114	16	US-10-751-736-99	Sequence 99, Appl
42	132.5	3.4	1128	17	US-10-732-923-3317	Sequence 3317, Ap
43	132.5	3.4	2265	15	US-10-092-900A-296	Sequence 296, App
44	132.5	3.4	2327	15	US-10-092-900A-294	Sequence 294, App
45	131.5	3.4	722	16	US-10-437-963-176171	Sequence 176171,

```
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 40109
; LENGTH: 62
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC004148.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 7.3
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 6.6
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 6.3
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 7.8
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 6.5
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 5.3
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 7.9
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 7.4
; OTHER INFORMATION: EST_HUMAN HIT: B155230.1, EVALUE 1.00e-32
; OTHER INFORMATION: SWISSPROT HIT: O09175, EVALUE 1.50e+00
; US-09-864-761-40109

Query Match 7.3%; Score 285; DB 9; Length 62;
Best Local Similarity 85.5%; Pred. No. 4e-15;
Matches 53; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 287 PGNIKAVGSIHAASAAEDNYGDACVLCPCPNILVATESGMLYHCVVLGESEDD 346
|||||: : |||||||
Db 1 PGNIGKLGPLPMHPAAEDNYGDACVLCPCPNILVATESGMLYHCVVLGESEDD 60

QY 347 HT 348
||
Db 61 HT 62

RESULT 2
US-09-864-761-39809
; Sequence 39809, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
```

```
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 39809
; LENGTH: 40
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC004148.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.8
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.6
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.7
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.7
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.4
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.9
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.4
; OTHER INFORMATION: EST_HUMAN HIT: AA488609.1, EVALUE 1.00e-15
; US-09-864-761-39809

Query Match 5.2%; Score 201; DB 9; Length 40;
Best Local Similarity 100.0%; Pred. No. 1.1e-08;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 550 SEKDIAAPPECLQLLSRATQVFREQYILKQDLAKEEIOR 589
|||||: : |||||||
Db 1 SEKDIAAPPECLQLLSRATQVFREQYILKQDLAKEEIOR 40

RESULT 3
US-10-437-963-140355
; Sequence 140355, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2005, 18:02:36 ; Search time 44 Seconds  
(without alignments)  
1257.158 Million cell updates/sec

Title: US-09-684-890A-2  
Perfect score: 3887  
Sequence: 1 MAAAEQVGGELWQTLVFN.....GEHREKVKQINDIRHVNFP 741

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3817	98.2	745	4	US-09-949-016-9976
2	156	4.0	192	4	US-09-270-767-44575
3	149.5	3.8	2954	4	US-09-150-867-1
4	142.5	3.7	2662	4	US-09-595-684B-31
5	142.5	3.7	2663	4	US-09-538-092-1252
6	137.5	3.5	1780	4	US-09-949-016-6899
7	137.5	3.5	1786	4	US-09-949-016-7880
8	134	3.4	606	4	US-08-477-831C-2
9	134	3.4	1979	4	US-09-949-016-6468
10	134	3.4	2047	4	US-09-949-016-7404
11	134	3.4	2482	1	US-08-328-254-6
12	134	3.4	3210	4	US-09-538-092-1154
13	134	3.4	3248	1	US-08-353-700-1
14	134	3.4	3248	5	PCT-US95-16216-1
15	131.5	3.4	976	3	US-09-104-324B-4
16	131.5	3.4	976	4	US-09-538-092-1339
17	131	3.4	631	4	US-08-477-831C-11
18	128	3.3	514	2	US-08-960-022-14
19	127.5	3.3	1972	4	US-08-875-435B-3
20	126.5	3.3	1886	3	US-08-938-105-3
21	126	3.2	1055	4	US-09-949-016-9776
22	125.5	3.2	2285	3	US-09-308-375-2
23	125.5	3.2	2285	4	US-09-932-183A-2
24	125	3.2	630	4	US-09-248-796A-20275
25	125	3.2	1201	3	US-09-098-901-2
26	123	3.2	994	4	US-09-949-016-6779
27	122.5	3.2	592	2	US-08-736-770-6

28 122.5 3.2 592 4 US-09-702-705-1809 Sequence 1809, Ap  
29 122.5 3.2 592 4 US-09-736-457-1809 Sequence 1809, Ap  
30 122.5 3.2 592 4 US-09-643-657-4 Sequence 4, Appli  
31 122.5 3.2 592 4 US-09-671-325-1809 Sequence 1809, Ap  
32 122 3.1 1898 1 US-08-056-200-94 Sequence 94, Appl  
33 122 3.1 1898 2 US-08-800-644-94 Sequence 94, Appl  
34 122 3.1 1898 4 US-09-538-092-1280 Sequence 1280, Ap  
35 122 3.1 1972 4 US-08-875-435B-4 Sequence 4, Appli  
36 121.5 3.1 605 4 US-09-949-016-8823 Sequence 8823, Ap  
37 121.5 3.1 1799 4 US-09-134-000C-5178 Sequence 5178, Ap  
38 121 3.1 589 4 US-09-643-657-14 Sequence 14, Appl  
39 121 3.1 631 4 US-09-949-016-11595 Sequence 11595, A  
40 121 3.1 1786 3 US-08-973-462-8 Sequence 8, Appli  
41 120.5 3.1 1088 3 US-09-085-199B-11 Sequence 11, Appl  
42 120 3.1 2568 4 US-09-866-108A-3 Sequence 3, Appli  
43 119.5 3.1 1307 4 US-09-949-016-7561 Sequence 7561, Ap  
44 119.5 3.1 1401 4 US-09-750-590A-2 Sequence 2, Appli  
45 119.5 3.1 1857 4 US-09-917-254-91 Sequence 91, Appl

## ALIGNMENTS

## RESULT 1

US-09-949-016-9976  
; Sequence 9976, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 9976  
; LENGTH: 745  
; TYPE: PRT  
; ORGANISM: Human  
US-09-949-016-9976

Query Match 98.2%; Score 3817; DB 4; Length 745;  
Best Local Similarity 98.4%; Pred. No. 0;  
Matches 729; Conservative 3; Mismatches 9; Indels 0; Gaps 0;  
Qy 1 MAAAEQVGGELWQTLVFNHVVFLRLREGLNQSPTEAEKPASSSLPSSPPOLLTRNV 60  
Db 5 MAAAEQVGGELWQTLVFNHVVFLRLREGLNQSPTEAEKPASSSLPSSPPOLLTRNV 64  
Qy 61 VFGLGELFLWDGEDSSFLVRLRPGSGGEEPALSOYQLLCLINPPLFBIYQVLLSPTQ 120  
Db 65 VFGLGELFLWDGEDSSFLVRLRPGSGGEEPALSOYQLLCLINPPLFBIYQVLLSPTQ 124  
Qy 121 HHVALIGIKLWLELPKRWGKNSEFEGKSTVNCSTTPVAERFFTSSTSLTLKHAAYP 180  
Db 125 HHVALIGIKLWLELPKRWGKNSEFEGKSTVNCSTTPVAERFFTSSTSLTLKHAAYP 184  
Qy 181 SEILDHVVLLTSDNIRIYSLREPOTPTNVIILSEAEESVLNKGRAVTASIGETAVA 240  
Db 185 SEILDHVVLLTSDNIRIYSLREPOTPTNVIILSEAEESVLNKGRAVTASIGETAVA 244  
Qy 241 FDFGDLDAVPEKTLFGQNGKDEVVAYPLYIYENGETFLTYSLLHSPGNWKVAGVSIHAH 300  
Db 245 FDFGDLAAVPEKTLFGQNGKDEVVAYPLYIYENGETFLTYSLLHSPGNKGLGLPLPMH 304  
Qy 301 SAAEDNYGDACAVLCPLPCVFNILVIATSGMGLYHCVVLEGEDEDDHSTSEKSWDSRIDLI 360

```
Db 305 PAAEDNYGYDACAVLCLCPVNLIVLATSGMLYHCVLEGEEDDHTSEKSWDSRIDLI 364
QY 361 PSILYFECVELELALKLASGEDDPDFSCPKLHRDPKPSRVHCTHEAGVHSGLTM 420
Db 365 PSILYFECVELELALKLASGEDDPDFSCPKLHRDPKPSRVHCTHEAGVHSGLTM 424
QY 421 IHKLHKLFGSDDEKDSLOELSTEQKCFVEHILCTPLPCROPAPIRGFWIVDPDILGPTM 480
Db 425 IHKLHKLFGSDDEKDSLOELSTEQKCFVEHILCTPLPCROPAPIRGFWIVDPDILGPTM 484
QY 481 ICITSTYECLINPLISTVHPASPPLCTREDVEAESSLRVLAETPDSFEKHRSILQRS 540
Db 485 ICITSTYECLINPLISTVHPASPPLCTREDVEAESSLRVLAETPDSFEKHRSILQRS 544
QY 541 VANPAFLKASEKDIAAPPPEECQLLSRATQVREQVILKODLAKBEIQRVVKLDCQKK 600
Db 545 VANPAFLKASEKDIAAPPPEECQLLSRATQVREQVILKODLAKBEIQRVVKLDCQKK 604
QY 601 QLEDLSYCREERKSREMAERLADKYEEAKEQEDIMNRMKLLHSFHSSELVLSDSERD 660
Db 605 QLEDLSYCREERKSREMAERLADKYEEAKEQEDIMNRMKLLHSFHSSELVLSDSERD 664
QY 661 MKKELQILPDQLRHGLGNAIKQVTMKDYQOQKMEKVLSPKPTIILSAYORKICISILKE 720
Db 665 MKKELQILPDQLRHGLGNAIKQVTMKDYQOQKMEKVLSPKPTIILSAYORKICISILKE 724
QY 721 EGEHIREMVKQINDIRNHVNF 741
Db 725 EGEHIREMVKQINDIRNHVNF 745
```

## RESULT 2

```
US-09-270-767-44575
; Sequence 44575, Application US/09270767
; Patent No. 6703491
```

## GENERAL INFORMATION:

```
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 44575
; LENGTH: 192
```

```
; TYPE: PRT
```

```
; ORGANISM: Drosophila melanogaster
```

```
; OTHER INFORMATION: Xaa means any amino acid
```

```
US-09-270-767-44575
```

```
Query Match 4.0%; Score 156; DB 4; Length 192;
Best Local Similarity 27.4%; Pred. No. 3.5e-06;
Matches 46; Conservative 40; Mismatches 74; Indels 8; Gaps 4;
```

```
QY 574 EYLLKQDLAKBEIQRVVKLDCQKKQLEDLSYCREERKSREMAERLADKYEEAKEQ 633
Db 1 EYLLKQDLAKBEIQRVVKLDCQKKQLEDLSYCREERKSREMAERLADKYEEAKEQ 60
QY 634 EDIMNRMKLLHSFHSSELVLSDSERDKKELQLIPDQLRHGLGNAIKQVTMKDYQOQRM 693
Db 61 ELLVRKCNALMQRANSLPNSVIAEREFQSVIRLNKVTQSLAAGLE--TAKTFNKQRY 118
QY 694 EKVLG---LPKPTIILSAYORKICISILKE--EGEHIEM--VKQINDI 735
Db 119 HIAQSOEDLKNAYELPERQHRITTEILTLQAGEIDRQITDKRINKI 166
```

## RESULT 3

```
US-09-150-867-1
```

```
; Sequence 1, Application US/09150867
```

```
; Patent No. 6645748
```

```
; GENERAL INFORMATION:
; APPLICANT: Wood, Kenneth W.
; APPLICANT: Sakowicz, Roman
; APPLICANT: Goldstein, Lawrence S.B.
; APPLICANT: Cleveland, Don W.
; TITLE OF INVENTION: The Regents of the University of California
; TITLE OF INVENTION: Plus End-Directed Microtubule Motor Required for
; FILE REFERENCE: 18557C-000110US
; CURRENT APPLICATION NUMBER: US/09/150,867
; CURRENT FILING DATE: 1998-09-10
; EARLIER APPLICATION NUMBER: US 60/058,645
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2954
; TYPE: PRT
; ORGANISM: Xenopus sp.
; FEATURE:
; OTHER INFORMATION: Xenopus centromere-associated protein-E (XCENP-E)
; OTHER INFORMATION: member of the kinesin superfamily of microtubule
; OTHER INFORMATION: motor proteins
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (1)..(472)
; OTHER INFORMATION: kinesin like motor domain
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (473)..(2752)
; OTHER INFORMATION: rod domain
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (2753)..(2954)
; OTHER INFORMATION: tail domain
; OTHER INFORMATION: tail domain
US-09-150-867-1
```

```
Query Match 3.8%; Score 149.5; DB 4; Length 2954;
Best Local Similarity 18.5%; Pred. No. 0.0013;
Matches 74; Conservative 81; Mismatches 157; Indels 89; Gaps 14;
```

```
QY 370 EUELAKLKGASDEDDPFDSCFVKLHRDPKPSRVHCTHEAGVHSGLTHLKHKLFG 429
Db 1581 ELQAKNLALIAASD-----NCPITQEKETSA----DCVHPLEEKILLIT--EELHQTN 1628
QY 430 SDEE---DKDSLQELSTEQKCFVEHILCTPLPCROPAPIRGFWIVDPDILGPTMCIITST 486
Db 1629 EQEKLHKEKLEQAQVELKCEVHLM-----KSMIESKSS 1664
QY 487 YECLIWPLLLSTVHPASPPLCTREDVEAESSLRVLAETPDSFEKHRSI-----LQRSV 541
Db 1665 LE-----SLQHEKHDTEQQLLALKQOQVVOYQEKELQOQTHEULTAEVDHLKENIELGLNF 1720
QY 542 ANPAFLKASEKDIAAPPPEECQLLSRATQVREQVILKODLAKBEIQRVVKLDCQKKQ 601
Db 1721 KNEAQOQTKT-----EQC--LINENKELEQSOHRLQCEI--EELMKSLEK--DKESA 1765
QY 602 LEDLSYCREERKSREMAERLADKYEEAKEQEDIMNRMKLLHSFHSSELVLSDSERDM 661
Db 1766 LETLKESEQKVINLQEMEMVMELEBKNSQRTVIAERDQLODDLRESEVMSIETQDDL 1825
QY 662 KK-----ELQLIPDQLRHGLGNAIKQVTMKDYQOQKMEKVL 697
Db 1826 RKAQELAQOQKQVQELTSQISVLOEKISLLENQMLYNAVATVETLSERDDLQNSQKHFLF 1885
QY 698 SLPKPTIILSAYORK-CIQSILKEGEHIREMVKQINDIRN 737
Db 1886 S-EIETLSLSLKEFALEQAQEKDAARAKTIDITEKISN 1925
```

## RESULT 4

```
US-09-595-684B-31
```

```
; Sequence 31, Application US/09595684B
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2005, 17:47:00 ; Search time 169 Seconds  
(without alignments)  
1695.796 Million cell updates/sec

Title: US-09-684-890A-2

Perfect score: 3887

Sequence: 1 MAAGGPGVGDGLMQLP.....GSHIREMVQINDIRNVNF 741

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_16Dec04:\*

1: Geneseqp1980s:\*

2: Geneseqp1990s:\*

3: Geneseqp2000s:\*

4: Geneseqp2001s:\*

5: Geneseqp2002s:\*

6: Geneseqp2003as:\*

7: Geneseqp2003bs:\*

8: Geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3887	100.0	741	AAE01161	AAE01161 Human Nup
2	3874	99.7	741	AAAM39226	AAAM39226 Human pol
3	3871	99.6	757	AAAM79827	AAAM79827 Human pro
4	3871	99.6	757	AAAM79826	AAAM79826 Human pro
5	3871	99.6	757	AAAM41012	AAAM41012 Human pol
6	3817	98.2	741	ADK60214	ADK60214 Angiogene
7	3817	98.2	741	ADK60515	ADK60515 Angiogene
8	3817	98.2	741	ADP71138	ADP71138 Angiogene
9	3817	98.2	741	ABM61818	ABM61818 Tumour-as
10	3461.5	89.1	742	AAE01162	AAE01162 Rat Nup88
11	2313	59.5	443	AAAM78842	AAAM78842 Human pro
12	1561	40.2	303	AAAM78843	AAAM78843 Human pro
13	648.5	16.7	702	AAE01160	AAE01160 Drosophil
14	646.5	16.6	702	ABBT71509	ABBT71509 Drosophil
15	285	7.3	62	AAAM20196	AAAM20196 Peptide #
16	285	7.3	62	ABBA40469	ABBA40469 Peptide #
17	285	7.3	62	AAAM34190	AAAM34190 Peptide #
18	285	7.3	62	ABB24811	ABB24811 Protein #
19	285	7.3	62	AAAM74021	AAAM74021 Human bon
20	285	7.3	62	AAAM61293	AAAM61293 Human bon
21	285	7.3	62	ABGS5775	ABGS5775 Human liv
22	285	7.3	62	ABG43915	ABG43915 Human liv
23	245.5	6.3	822	AAAG41929	AAAG41929 Arabidops
24	219.5	5.6	767	AAAG41930	AAAG41930 Arabidops
25	211.5	5.4	725	AAAG41931	AAAG41931 Arabidops

#### ALIGNMENTS

##### RESULT 1

AAE01161

ID AAE01161 standard; protein; 741 AA.

XX AAE01161;

XX 17-JUL-2001 (first entry)

XX Human Nup88 homology protein.

XX Transcription factor-selective nuclear transport receptor; dorsal;

KW nuclear factor-kappa B; NF-kB; nuclear pore; humoral response; therapy;

KW inflammation; asthma; rheumatoid arthritis; RA; septic shock;

KW chronic obstructive pulmonary disease; COPD; angiogenesis; lung fibrosis;

KW glomerulonephritis; atherosclerosis; apoptotic disorder; antibacterial;

KW acquired immune deficiency syndrome; AIDS; immunosuppressive;

KW nephrotropic; human; Nup88.

XX Homo sapiens.

XX WO200129087-A1.

XX 26-APR-2001.

XX 19-OCT-2000; 2000WO-SE002022.

XX 22-OCT-1999; 99SE-00003832.

XX (INNA-) INNATE PHARM AB.

XX Samakovlis C, Uv AE;

XX WPI; 2001-290899/30.

XX Novel purified transcription factor-selective nuclear transport receptor

PT polypeptide is used to treat, prevent and diagnose inflammation, asthma,

PT rheumatoid arthritis, atherosclerosis, AIDS glomerulonephritis and

PT apoptotic disorders.

XX Claim 6; Page 35-37; 42pp; English.

XX The present sequence is human Nup88 protein which localise to the

CC cytoplasmic filaments of the nuclear pores to provide binding sites for

CC nuclear import substrates. The Nup88 protein has homology to Drosophila

CC transcription factor-selective nuclear transport receptor. The

CC transcription factor-selective nuclear transport receptor mediates the

CC import of dorsal nuclear factor-kappa B (NF-kB) protein at the level of

CC the nuclear pore and is required for activation of the Drosophila humoral

AAE01161 Peptide #  
AAE01161 Peptide #  
AAE01161 Peptide #  
AAE01161 Protein #  
AAE01161 Human bon  
AAE01161 Human bra  
AAE01161 Human liv  
AAE01161 Human pep  
AAE01161 Angiogene  
AAE01161 Antipsoori  
AAE01161 PRO poly  
AAE01161 Human pro  
AAE01161 Human pro  
AAE01161 Human pro  
AAE01161 Human ret  
AAE01161 Human pro  
AAE01161 Hyperther  
AAE01161 Amino aci  
AAE01161 Transport

CC immune response. It is also used for prevention, diagnosis and treatment  
CC of pathophysiological disorders related to the family of nuclear  
CC receptors such as inflammation, asthma, rheumatoid arthritis (RA),  
CC chronic obstructive pulmonary disease (COPD), angiogenesis, septic shock,  
CC lung fibrosis, glomerulonephritis, atherosclerosis, AIDS and apoptotic  
CC disorders. It is also used in screening assays to identify its agonists  
CC or inhibitors  
XX  
SQ Sequence 741 AA;

Query Match 100.0%; Score 3887; DB 4; Length 741;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 741; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MAAAGPVGDELWQTLNHNHVFRLRGLKQSPTEAEKPASSSLPSPPPQLLRNV 60  
DB 1 MAAAGPVGDELWQTLNHNHVFRLRGLKQSPTEAEKPASSSLPSPPPQLLRNV 60  
QY 61 VFGLGELFLWDGEDSSFLVRLRGPSSGGEPPALSYQVRLCINPPLFEIYQVLLSPTQ 120  
DB 61 VFGLGELFLWDGEDSSFLVRLRGPSSGGEPPALSYQVRLCINPPLFEIYQVLLSPTQ 120  
QY 121 HHVALIGIKGLMWLELPKRWGNSEPEGGKSTVNCSTTPVAERFFTSSTSLTLKHAAYTP 180  
DB 121 HHVALIGIKGLMWLELPKRWGNSEPEGGKSTVNCSTTPVAERFFTSSTSLTLKHAAYTP 180  
QY 181 SETLDPHVLLTSDNVIRIYSLREPOTPNVILSEAEESVLNKGRAVYASLGATA 240  
DB 181 SETLDPHVLLTSDNVIRIYSLREPOTPNVILSEAEESVLNKGRAVYASLGATA 240  
QY 241 FDFGGLDAPVKTFLFGQNGDEVVAYPLYLYENGETFLTYISLLHSPGNIKAVGSIHA 300  
DB 241 FDFGGLDAPVKTFLFGQNGDEVVAYPLYLYENGETFLTYISLLHSPGNIKAVGSIHA 300  
QY 301 SAAEDNYGYDACAVLCLCPVNLIVTATSGMLYHCVLEGEEDDHTSEKSWDSRIDLI 360  
DB 301 SAAEDNYGYDACAVLCLCPVNLIVTATSGMLYHCVLEGEEDDHTSEKSWDSRIDLI 360  
QY 361 PSLYVFECELELALKLAGEDDPDSFSCPKLHRDPCPSRVHCTHEAGVHSLGTW 420  
DB 361 PSLYVFECELELALKLAGEDDPDSFSCPKLHRDPCPSRVHCTHEAGVHSLGTW 420  
QY 421 IHKLHKLGSDEEDKDSLGELSTEQCFVEHILCTPLPCROPAPTRGFVWIPDILGPTM 480  
DB 421 IHKLHKLGSDEEDKDSLGELSTEQCFVEHILCTPLPCROPAPTRGFVWIPDILGPTM 480  
QY 481 ICITSTYECLIWPLSTVHPASPLLCITREDVEVAESSLRVLAETPDSFEKHRSILQRS 540  
DB 481 ICITSTYECLIWPLSTVHPASPLLCITREDVEVAESSLRVLAETPDSFEKHRSILQRS 540  
QY 541 VANPAFLKASEKDIAAPPPECQLLSRATQVFEQVILKODIAKEIQRVVKLLCDQKK 600  
DB 541 VANPAFLKASEKDIAAPPPECQLLSRATQVFEQVILKODIAKEIQRVVKLLCDQKK 600  
QY 601 QLEDLSYCREERSKREMAERLADKVEEAKQEDIMNMKKLLSHFHSSELVLSDSER 660  
DB 601 QLEDLSYCREERSKREMAERLADKVEEAKQEDIMNMKKLLSHFHSSELVLSDSER 660  
QY 661 MKKELQILPDQLRHGNAIKQVTMKDYQOQKMEKVLSPKPTIILSAVQRKCIQSILKE 720  
DB 661 MKKELQILPDQLRHGNAIKQVTMKDYQOQKMEKVLSPKPTIILSAVQRKCIQSILKE 720  
QY 721 EGEHIREMVKQINDIRNVNF 741  
DB 721 EGEHIREMVKQINDIRNVNF 741

RESULT 2  
ID AAM39226  
AC AAM39226 standard; protein; 741 AA.  
XX AAM39226;  
XX

DT 22-OCT-2001 (first entry)  
XX Human polypeptide SEQ ID NO 2371.  
XX  
XX Human; nototropic; immunosuppressant; cytostatic; gene therapy; cancer;  
KW peripheral nervous system; neuropathy; central nervous system; CNS;  
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
KW leukaemia.  
XX  
XX Homo sapiens.  
OS WO200153312-A1.  
PN  
XX 26-JUL-2001.  
PD  
XX  
PF 26-DEC-2000; 2000WO-US034263.  
XX  
XX 23-DEC-1999; 99US-00471275.  
PR 21-JAN-2000; 2000US-00488725.  
PR 25-APR-2000; 2000US-00552317.  
PR 20-JUN-2000; 2000US-00598042.  
PR 19-JUL-2000; 2000US-00620312.  
PR 03-AUG-2000; 2000US-00653450.  
PR 14-SEP-2000; 2000US-00662191.  
PR 19-OCT-2000; 2000US-00693036.  
PR 29-NOV-2000; 2000US-00727344.  
XX  
(HYSE-) HYSEQ INC.  
XX  
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;  
PI Zhou P, Goodrich R, Drmanac RT;  
XX  
XX WPI; 2001-442253/47.  
DR N-PSDB; AAI58382.  
XX  
XX Novel nucleic acids and polypeptides, useful for treating disorders such  
PT as central nervous system injuries.  
XX  
XX Example 4; SEQ ID NO 2371; 10078pp; English.  
XX  
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and the  
CC encoded polypeptides (AAM38642-AAM42213) with nototropic,  
CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
CC in gene therapy. A composition containing a polypeptide or polynucleotide  
CC of the invention may be used to treat diseases of the peripheral nervous  
CC system, such as peripheral nervous injuries, peripheral neuropathy and  
CC localised neuropathies and central nervous system diseases, such as  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities such as: Immune system suppression,  
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukaemias and  
CC C.N.S disorders. Note: The sequence data for this patent did not form  
CC part of the printed specification  
XX  
SQ Sequence 741 AA;

Query Match 99.7%; Score 3874; DB 4; Length 741;  
Best Local Similarity 99.7%; Pred. No. 0;  
Matches 739; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 MAAAGPVGDELWQTLNHNHVFRLRGLKQSPTEAEKPASSSLPSPPPQLLRNV 60  
DB 1 MAAAGPVGDELWQTLNHNHVFRLRGLKQSPTEAEKPASSSLPSPPPQLLRNV 60  
QY 61 VFGLGELFLWDGEDSSFLVRLRGPSSGGEPPALSYQVRLCINPPLFEIYQVLLSPTQ 120  
DB 61 VFGLGELFLWDGEDSSFLVRLRGPSSGGEPPALSYQVRLCINPPLFEIYQVLLSPTQ 120

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 20, 2005, 06:10:34 ; Search time 7720 Seconds  
(without alignments)  
11724.978 Million cell updates/sec

Title: US-09-684-890A-1  
Perfect score: 2378  
Sequence: 1 gataaacccacagacacaa.....tataaaagggttttgatg 2378

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:\*

1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_hic:\*  
4: gb\_est3:\*  
5: gb\_est4:\*  
6: gb\_est5:\*  
7: gb\_est6:\*  
8: gb\_gss1:\*  
9: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1685	70.9	2430	3 AK02643	AK02643 Mus muscu
2	1671.2	70.3	2452	3 AK028563	AK028563 Mus muscu
3	1552.6	65.3	2076	3 AK085776	AK085776 Mus muscu
4	1220.8	51.3	3004	3 AK050893	AK050893 Mus muscu
5	1014.8	42.7	1978	3 AK039007	AK039007 Mus muscu
c	6	918.8	38.6	954 1	AL525959
7	894.6	37.6	935 1	AL526005	AL526005 AL526005
8	865.6	36.4	947 5	BU193560	BU193560 AGENCOURT
9	859.6	36.1	1096 5	BM923340	BM923340 AGENCOURT
10	855	36.0	1149 4	BM541613	BM541613 AGENCOURT
11	837	35.2	921 5	BU526916	BU526916 AGENCOURT
c	12	823	34.6	932 5	BX349939
13	820.2	34.5	897 5	BX349940	BX349940 AGENCOURT
c	14	780.8	32.8	945 5	BQ953718
15	778	32.7	887 6	CD359338	CD359338 AGENCOURT
16	769.2	32.3	843 4	BG759661	BG759661 602713354
17	764	32.1	866 5	BUB58009	BUB58009 AGENCOURT
18	762.2	32.1	902 5	BU539430	BU539430 AGENCOURT
19	761.2	32.0	789 4	B1819253	B1819253 603034680
20	758	31.9	921 5	BQ647182	BQ647182 AGENCOURT
21	755.2	31.8	833 4	B1552947	B1552947 603197875
22	742.4	31.2	873 4	BG219184	BG219184 RST38937
23	737.6	31.0	762 4	BG686994	BG686994 602650990
24	735.2	30.9	795 6	CB997327	CB997327 AGENCOURT

#### ALIGNMENTS

RESULT 1  
AK002643 2430 bp mRNA linear HTC 03-APR-2004  
LOCUS Mus musculus adult male kidney cDNA, RIKEN full-length enriched library, clone:0610016P04 product:preimplantation protein 2, full insert sequence.  
DEFINITION AK002643  
ACCESSION AK002643.1 GI:112832779  
VERSION HTc; CAP trapper.  
KEYWORDS Mus musculus (house mouse)  
SOURCE Mus musculus  
ORGANISM Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1  
AUTHORS Carninci, P. and Hayashizaki, Y.  
TITLE High-efficiency full-length cDNA cloning  
JOURNAL Meth. Enzymol. 303, 19-44 (1999)  
MEDLINE 99279253  
PUBMED 10349636  
REFERENCE 2  
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)  
MEDLINE 20499374  
PUBMED 11042159  
REFERENCE 3  
AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsumura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer  
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)  
MEDLINE 20530913  
PUBMED 11076861  
REFERENCE 4  
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.  
TITLE Functional annotation of a full-length mouse cDNA collection  
JOURNAL Nature 409, 685-690 (2001)  
PUBMED 11076861  
REFERENCE 5  
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research

25 734.8 30.9 816 4 B1549458  
26 733.4 30.8 832 4 B1461310  
27 732.6 30.8 831 5 B1558277  
c 28 727 30.6 807 4 B194807  
29 719.4 30.3 799 2 BE902870  
30 717.6 30.2 740 7 CF122933  
31 715.2 30.1 796 4 B187429  
c 32 709 29.8 752 5 BQ182096  
c 33 704.8 29.6 756 1 AA868783  
34 700.2 29.4 909 6 CD557705  
35 698 29.4 816 4 B1458058  
36 697.8 29.3 784 4 BG686048  
37 697.8 29.3 792 6 CB990577  
38 697.4 29.3 839 4 B1598624  
39 693 29.1 718 4 BG720209  
40 692.6 29.1 798 4 B1550338  
41 690.6 29.0 832 5 BU940824  
42 682.8 28.7 783 4 B1459146  
43 680.4 28.6 690 4 BG719270  
44 677.4 28.5 730 4 BG718589  
45 673.8 28.3 965 4 B1461277

Group Phase I & II Team.  
 Analysis of the mouse transcriptome based on functional annotation  
 of 60,770 full-length cDNAs  
 Nature 420, 563-573 (2002)  
 6 (bases 1 to 2430)  
 Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Arai, A., Aono, H.,  
 Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Fukunishi, Y.,  
 Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K.,  
 Hiraoka, T., Hori, F., Imorani, K., Ishii, Y., Itoh, M., Izawa, M.,  
 Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M.,  
 Koyu, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K.,  
 Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C.,  
 Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D.,  
 Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y.,  
 Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T.,  
 Tejima, Y., Toya, T., Yamamura, T., Yasunishi, A., Yoshida, K.,  
 Yoshino, M., Muramatsu, M. and Hayashizaki, Y.  
 Direct Submission  
 Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of  
 Physical and Chemical Research (RIKEN), Laboratory for Genome  
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
 RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,  
 Kanagawa 230-0045, Japan (E-mail: genome-res@gscl.riken.jp,  
 URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222,  
 Fax: 81-45-503-9216)  
 Please visit our web site (http://genome.gsc.riken.jp/) for further  
 details.  
 cDNA library was prepared and sequenced in Mouse Genome  
 Encyclopedic Project of Genome Exploration Research Group in Riken  
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
 Division of Experimental Animal Research in Riken contributed to  
 prepare mouse tissues. First strand cDNA was primed with a primer  
 [5', GAGAGAGAGCGCGCCCACTCGAGTTTCTTTTCTTNN 3'], cDNA was  
 prepared by using triethanolamine-activated reverse transcriptase  
 and subsequently enriched for full-length by cap-trapper. cDNA went  
 through one round of normalization to Rot = 5.0. Second strand cDNA  
 was prepared with the primer adapter of sequence [5',  
 GAGAGAGAGAGTCCCAAGCTCAATTAATTAATTAACCCCGCCCG 3']. cDNA was  
 cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end:  
 XhoI. Host: SOUR.  
 Location/Qualifiers  
 1..2430  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="PANTOM DB:0610016P04"  
 /db\_xref="taxon:10090"  
 /clone="0610016P04"  
 /sex="male"  
 /tissue\_type="kidney"  
 /clone\_lib="RIKEN full-length enriched mouse cDNA library"  
 /dev\_stage="adult"  
 misc\_feature  
 10..2252  
 /note="preimplantation protein 2 (MGD|GI:104900,  
 GI|AK002643, evidence: BLASTN, 99%, match=1862)  
 putative"  
 polyA\_signal  
 2416..2421  
 /note="putative"  
 polyA\_site  
 2430  
 /note="putative"  
 ORIGIN  
 Query Match 70.9%; Score 1685; DB 3; Length 2430;  
 Best Local Similarity 86.0%; Pred. No. 0;  
 Matches 1946; Conservative 0; Mismatches 290; Indels 27; Gaps 6;  
 45 CAAGATGGCGCGCGCGAGGACCGGTGGCGGCGAGCTGTGGCAGACCTGGCTTCC 104  
 |||||  
 6 CAAGATGGCGCTGCCGTGGGCGCTTGGCGATGGGAGCTGTGCAGAGCTGGCTTCC 65  
 |||||  
 105 TAAACACGTCGTGTTCTTGGGCTCCGGAGGGAGCTGAATAAACAGAGTCCACCGAAGC 164  
 |||||  
 66 TAAACACGTCGTGTTCTTGGGCTCCGGCGGCGTGAAGAAACAGAGTCCAGCGGAGC 125  
 |||||

165 TGAGAAACAGCTTCTTCTGCTGTTGCTT-----CGTCGCGCGCGCGCGAGTTGCTGAC 218  
 |||||  
 126 GGAGAAAGCAGCGGCTTCGACTTCACCTCGTGGCGCGCTCTCTGCGCGCGGATTTGCGGAC 195  
 |||||  
 219 GAGAAACGTGGTCTTTTGGCTCGCGGAGAGCTTTTCTGTGGGCGGAGAGACAGCTC 278  
 |||||  
 186 GAGAAACCTGGTCTTCGCGCTCGGAGGGGAACTGTTTCTGTGGGAC-GGGAAGGACGCGC 244  
 |||||  
 279 CTTCTTAGTCTGCTCGCTTGGGCGGCGCGCGGCGGAGAGCGCGCGCTGTCCTCCCA 338  
 |||||  
 245 CTTCTTGGTGGTTCGCTTTCGAGGCGCGCGGCTGGTGGCGGAGCGCTCTCTCTCCCA 304  
 |||||  
 339 GTACACAGAGATGCTTTTGTGATAAATCCACCCCTGTTTGAATCTATCAAGTCTTGTGAAG 398  
 |||||  
 305 GTATCAGAGATTACTCTGCAATTAATCCACCCCTGTTTGAATTCATCAAGCTCTTGTGAG 364  
 |||||  
 399 CCCAACACACATCATGTAGCACTTATAGGAATAAAGGACTTATGGTATTAGAATTACC 458  
 |||||  
 365 TCCAAACCAACATCATGTAGCACTTATCGAAGTAAAGGACTTATGGCATTTAGAAATTACC 424  
 |||||  
 459 TAAAGATGGGGAAGAATCTGAATTTCAAGGTGGAATAATCAACAGTGAATTTGTATAC 518  
 |||||  
 425 TCAGAGGTGGGGAAGGACTCTGAATTTGAAGGTGGAAAGCAACTGTGAATTTGTATGAC 484  
 |||||  
 519 CACTCCAGTTCGCGAGAGATTTTTCACCAAGTTCCACCTCTCTGACTCTTAAAGCATGCTGC 578  
 |||||  
 485 CATCCGATGCTGAGAGATTTTTCACCAAGCTCTTACCTCTCTGACTCTGAGCATGCTGC 544  
 |||||  
 579 ATGGTATCAAGTGAATCTCGATCCCACTAGTGTGTTTAAACATCAGACAACTAAT 638  
 |||||  
 545 ATGGTATCAAGTGAATGCTGATGCTGCTCCCACTAGTGTGTTTGAACATCAGACAACTGAT 604  
 |||||  
 639 CAGAAATTTACTCTTACGTGAGCGGAGACACCCACTAAGCTGATTAATTTTCAAGAC 698  
 |||||  
 605 AAGAAATTTATTTCTCTCGGTGAGCGGCGGAGACACCCCACTAAGGTGATTTGTACTTTCA 664  
 |||||  
 699 CGAAGAGGAAAGTCTAGTACTCAATAAAGGAAGGCGGTATACCGCATCTCTTAGGAGAGAC 758  
 |||||  
 665 AGAAGAGGAAAGTTTAATCTCAATAAAGGAAGCATATACAGCTCTCTTAGGAGAGAC 724  
 |||||  
 759 AGCAGTTGCAATTTGATTTTGGGCCATTGGACCGCAGTCCCAAGACTCTATTATGGACAAA 818  
 |||||  
 725 TGCAGTGGCGTTGACTTCGCGGCCCTGCTAAACCGTCTCAAAGAAATATATTGAAACAGAA 784  
 |||||  
 819 CGGCAAGATGAAGTAGTGGCATACCACTGTACATTTATATGAAATGGAGAGACTTT 878  
 |||||  
 785 AGACAGAGAT---GTGGTGGCGGTATCCACTGTACATCTCTGTATGAGAAATGGGAGACCTT 841  
 |||||  
 879 CTTGACATACATCAGTCTGTACACAGCCCTCGAAATATTTGAAAGCTGTTGGGTCCAT 938  
 |||||  
 842 CTTACTACGTGAGGCTGTATACAGCCCGGGAATA-TTGGGAAGCTGTTGGGCGCCAC 900  
 |||||  
 939 TGCCCATGCAT-CTCGGCTGAAGATACTATGGTTATGATCGGTGTGCTGACTCTGCT 997  
 |||||  
 901 TGCCTATGCATCTCGCAGCTGAAGATACTACGGTTACGATGCTGTGCTACTACTCTGTT 960  
 |||||  
 998 TACCTGTGTCCCAATATCTTAGTGTGCTACTGAAATCAGAAATGCTGTATCACTGTG 1057  
 |||||  
 961 TGCCCTGTGTTCCAAATATCTTAGTAATTTGCGACTGAGTCAGGAATGCTGTACCCTGTG 1020  
 |||||  
 1058 TCGTGTAGAAAGGGAAGAAAGATGACCAACACCGTCAGAAAAGCTCTGGGATTCAGGA 1117  
 |||||  
 1021 TTGTACTAGAGGAGAGAGAAAGATGACCAACCGTTAGAAAAGTCTGGGATCCAGGG 1080  
 |||||  
 1118 TTGACCTCATCTCTCTGTATGTGTTTGAATGTGTTGAGTTGGAGCTGTGTTGAAAC 1177  
 |||||  
 1081 CTGACTTCATCCCTCTCTGTACGTGTTTGAAGTGTGTTGAGTTAGAGCTTGGCTGGAAGC 1140  
 |||||  
 1178 TGGCATCTGGAGAGATGACCCCTTTGATTTCTGACTTTTCTTGTTCAGTCAAACTCAT 1237  
 |||||  
 1141 TGGCATCTGGAGAGAGCGATCCCTTTGCTCTGACTTTTCTTCTGCGCAATTAACATGACA 1200  
 |||||

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 20, 2005, 11:18:28 ; Search time 1482 Seconds

(without alignments)  
10179.925 Million cell updates/sec

Title: US-09-684-890A-1

Perfect score: 2378

Sequence: 1 gataaccacacagacacaa.....tataaaagggtgtttgatg 2378

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 7173243 seqs, 3172129809 residues

Total number of hits satisfying chosen parameters: 14346486

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA.\*

1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq.\*  
10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq.\*  
12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*  
13: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq.\*  
14: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq.\*  
15: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq.\*  
16: /cgn2\_6/ptodata/2/pubpna/US10D\_PUBCOMB.seq.\*  
17: /cgn2\_6/ptodata/2/pubpna/US10E\_PUBCOMB.seq.\*  
18: /cgn2\_6/ptodata/2/pubpna/US10F\_PUBCOMB.seq.\*  
19: /cgn2\_6/ptodata/2/pubpna/US10G\_PUBCOMB.seq.\*  
20: /cgn2\_6/ptodata/2/pubpna/US10H\_PUBCOMB.seq.\*  
21: /cgn2\_6/ptodata/2/pubpna/US10I\_PUBCOMB.seq.\*  
22: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*  
23: /cgn2\_6/ptodata/2/pubpna/US11A\_PUBCOMB.seq.\*  
24: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq.\*  
25: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*  
26: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2341.6	98.5	2392	15	US-10-037-270-261 Sequence 261, App
2	2341.6	98.5	2392	17	US-10-117-722-261 Sequence 261, App
3	1710.2	71.9	2229	18	US-10-261-175A-18 Sequence 18, Appl
c	1411.6	59.4	8263	17	US-10-062-674-1762 Sequence 1762, Ap
5	393.4	16.5	456	10	US-09-918-995-15834 Sequence 15834, A
6	332	14.0	356	9	US-09-796-692-2684 Sequence 2684, Ap
7	332	14.0	356	14	US-10-040-862-2684 Sequence 2684, Ap

8	332	14.0	356	17	US-10-057-475B-2684	Sequence 2684, Ap
9	332	14.0	356	17	US-10-154-884B-2684	Sequence 2684, Ap
10	332	14.0	356	19	US-10-764-324-2684	Sequence 2684, Ap
11	322.2	13.5	429	10	US-09-918-995-8447	Sequence 8447, Ap
12	243.4	10.2	514	16	US-10-029-386-2600	Sequence 2600, Ap
13	182.8	7.7	493	9	US-09-783-590-9025	Sequence 9025, Ap
C 14	166	7.0	466	9	US-09-864-761-6544	Sequence 6544, Ap
C 15	165.4	7.0	187	9	US-09-864-761-23256	Sequence 23256, A
C 16	146.4	6.2	212	9	US-09-933-797-642	Sequence 642, App
17	145	6.1	147	16	US-10-029-386-16300	Sequence 16300, A
C 18	134.6	5.7	459	9	US-09-864-761-6262	Sequence 6262, Ap
C 19	134.6	5.7	733	13	US-10-027-632-127047	Sequence 127047, Ap
C 20	134.6	5.7	733	17	US-10-027-632-127047	Sequence 127047, Ap
C 21	134.6	5.7	831	13	US-10-027-632-127046	Sequence 127046, Ap
C 22	134.6	5.7	831	17	US-10-027-632-127046	Sequence 127046, Ap
C 23	133.8	5.6	420	17	US-10-062-674-816	Sequence 816, App
C 24	122	5.1	122	9	US-09-864-761-22981	Sequence 22981, A
C 25	105.6	4.4	543	16	US-10-029-386-13044	Sequence 13044, A
C 26	103.4	4.3	112	16	US-10-029-386-26744	Sequence 26744, A
C 27	84	3.5	805	13	US-10-027-632-127043	Sequence 127043, A
C 28	84	3.5	805	13	US-10-027-632-127044	Sequence 127044, A
C 29	84	3.5	805	13	US-10-027-632-127045	Sequence 127045, A
C 30	84	3.5	805	13	US-10-027-632-154416	Sequence 154416, A
C 31	84	3.5	805	17	US-10-027-632-127043	Sequence 127043, A
C 32	84	3.5	805	17	US-10-027-632-127044	Sequence 127044, A
C 33	84	3.5	805	17	US-10-027-632-127045	Sequence 127045, A
C 34	84	3.5	805	17	US-10-027-632-154416	Sequence 154416, A
C 35	60	2.5	60	10	US-09-908-975-10796	Sequence 10796, A
C 36	54.4	2.3	65	10	US-09-908-975-27778	Sequence 27778, A
C 37	46	1.9	715	20	US-10-425-115-24320	Sequence 24320, A
C 38	45	1.9	1702	18	US-10-302-172-574	Sequence 574, App
C 39	44	1.9	694	14	US-10-184-644-60	Sequence 60, Appl
C 40	44	1.9	694	14	US-10-184-634-60	Sequence 60, Appl
C 41	42.6	1.8	414	19	US-10-437-963-56419	Sequence 56419, A
C 42	41.8	1.8	1840	17	US-10-264-049-284	Sequence 284, App
C 43	41.8	1.8	2383	17	US-10-104-047-1967	Sequence 1967, App
C 44	41.8	1.8	2383	21	US-10-887-553A-1206	Sequence 1206, App
C 45	41.2	1.7	145068	19	US-10-322-281-33	Sequence 33, Appl

#### ALIGNMENTS

#### RESULT 1

US-10-037-270-261  
; Sequence 261, Application US/10037270  
; Publication No. US20030104529A1  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Liu, Chenghua  
; APPLICANT: Asundi, Vinod  
; APPLICANT: Zhang, Jie  
; APPLICANT: Ren, Feiyang  
; APPLICANT: Chen, Rui-hong  
; APPLICANT: Zhao, Qing A.  
; APPLICANT: Wehrman, Tom  
; APPLICANT: Xue, Aidong J.  
; APPLICANT: Wang, Jian-Rui  
; APPLICANT: Zhou, Ping  
; APPLICANT: Ma, Yundong  
; APPLICANT: Wang, Dunrui  
; APPLICANT: Wang, Zhiwei  
; APPLICANT: Tillinghast, John  
; APPLICANT: Drmanac, Radoje T.  
; TITLE OF INVENTION: Polypeptides  
; FILE REFERENCE: 784CIP2B  
; CURRENT APPLICATION NUMBER: US/10/037,270  
; CURRENT FILING DATE: 2002-01-04  
; PRIOR APPLICATION NUMBER: 09/552,317  
; PRIOR FILING DATE: 2000-04-25  
; PRIOR APPLICATION NUMBER: 09/488,725

```

; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1104
; SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 261
; LENGTH: 2392
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (49)..(2274)
US-10-037-270-261

Query Match      98.5%; Score 2341.6; DB 15; Length 2392;
Best Local Similarity 99.2%; Pred. No. 0;
Matches 2353; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 7 CCCAAGACACAAACATACCTTTGAGCAGTGTGGCCCAAGATGGCGGCCCGGAGGGA 66
DB 7 CCCTGAGGCCAAGAAATTCGGCACGAGAGCAGTGTGGCCCAAGATGGCGGCCCGGAGGGA 66
QY 67 CCGGTGGCGGCGCGAGCTGTGGCAGACCTGGCTTCCTAAACACGTCGTGTTCTTGGG 126
DB 67 CCGGTGGCGGCGCGAGCTGTGGCAGACCTGGCTTCCTAAACACGTCGTGTTCTTGGG 126
QY 127 CTCGGGAGGAGCTGAAAAACAGAGTCCAAACCGAAGCTGAGAAACAGCTTCTTCGTGG 186
DB 127 CTCGGGAGGAGCTGAAAAACAGAGTCCAAACCGAAGCTGAGAAACAGCTTCTTCGTGG 186
QY 187 TTGCTTCTGTCGCGCGCGCGAGTTGCTGACGAGAAACAGTGGTCTTTGGCCGCGGA 246
DB 187 TTGCTTCTGTCGCGCGCGCGAGTTGCTGACGAGAAACAGTGGTCTTTGGCCGCGGA 246
QY 247 GAGCTTTTCTGTGGACGGAAGACAGCTCCTTCTTAGTGTGCTTCGCTTCGGGGCCCC 306
DB 247 GAGCTTTTCTGTGGACGGAAGACAGCTCCTTCTTAGTGTGCTTCGCTTCGGGGCCCC 306
QY 307 AGCGGCGGCGGGAAGAGCCGCCCTGTCCAGTACACAGAGATTGCTTTGCAATAATCCA 366
DB 307 AGCGGCGGCGGGAAGAGCCGCCCTGTCCAGTACACAGAGATTGCTTTGCAATAATCCA 366
QY 367 CCCCTGTTTGAATCTATCAAGTCTTGTGTTAAAGCCCAACACATCATGTAGCACTTATA 426
DB 367 CCCCTGTTTGAATCTATCAAGTCTTGTGTTAAAGCCCAACACATCATGTAGCACTTATA 426
QY 427 GGAATTAAGAGGACTTATGTTATTTAGAAATTTACCTTAAAGATGGGGAAGAAATTTCTGAA 486
DB 427 GGAATTAAGAGGACTTATGTTATTTAGAAATTTACCTTAAAGATGGGGAAGAAATTTCTGAA 486
QY 487 GAAAGTGGAAAAATCAACAGTGAATTTGATGACCTCCAGTTCGCGAGAGATTTTTCACC 546
DB 487 GAAAGTGGAAAAATCAACAGTGAATTTGATGACCTCCAGTTCGCGAGAGATTTTTCACC 546
QY 547 AGTTCCACCTCTGACTCTAAAGCATGTCGATGATGATGATGATGATGATGATGATGATGAT 606
DB 547 AGTTCCACCTCTGACTCTAAAGCATGTCGATGATGATGATGATGATGATGATGATGATGAT 606
QY 607 CACGTAGTCTTAAACATCAGAACCGTAATCAGAAATTTACTCTCTAGTCTAGTCTAGTCTAG 666
DB 607 CACGTAGTCTTAAACATCAGAACCGTAATCAGAAATTTACTCTCTAGTCTAGTCTAGTCTAG 666
QY 667 ACACCCACTTAACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 726
DB 667 ACACCCACTTAACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 726
QY 727 GGAAGGGGCTTACCGCATCTCTAGAGAGACAGCAGTTGCAATTTGATTTTGGGCGCATG 786
DB 727 GGAAGGGGCTTACCGCATCTCTAGAGAGACAGCAGTTGCAATTTGATTTTGGGCGCATG 786
QY 787 GACGAGTCCCAAGACCTTATTTGACAAACCGCAAGATGAAAGTGTGATGATGATGATGATGAT 846
DB 787 GACGAGTCCCAAGACCTTATTTGACAAACCGCAAGATGAAAGTGTGATGATGATGATGATGAT 846
QY 847 CTGTACATCTTATGAAAAATGGAGAGACTTTTCTGACATACATCAGTCTGTTACACAGC 906

```

```

|||||
847 CTGTACATCTTATATGAAAAATGGAGAGACTTTCCTGCATACATCAGTCTGTTACACAGC 906
QY |||||
907 CCTGGAATATTTTGGAAAGCTGTTCGGTCCATTCGCCATGCATCTCGCGCTGAAGATAAC 966
DB |||||
907 CCTGGAATATTTTGGAAAGCTGTTCGGTCCATTCGCCATGCATCTCGCGCTGAAGATAAC 966
QY |||||
967 TATGTTATGATGCGTGTGCTGTACTCTGCTTACCTGTGTGCCCAATCTTAGTGATC 1026
DB |||||
967 TATGTTATGATGCGTGTGCTGTACTCTGCTTACCTGTGTGCCCAATCTTAGTGATC 1026
QY |||||
1027 GCTACTGAATCAGGAATGCTGTATCACTGTGCTGTAGAGGGGAGAAAGAGATGAC 1086
DB |||||
1027 GCTACTGAATCAGGAATGCTGTATCACTGTGCTGTAGAGGGGAGAAAGAGATGAC 1086
QY |||||
1087 CACACGTGAGAAAAAGTCTGGGATTCAGGATTTGACCTCAATTCCTTCTGTGTGTTT 1146
DB |||||
1087 CACACGTGAGAAAAAGTCTGGGATTCAGGATTTGACCTCAATTCCTTCTGTGTGTTT 1146
QY |||||
1147 GAATGTGTGAGTTCGGAGCTTCTTTGAAACTTGGCATCTGGAGAGATGACCTTTTGAT 1206
DB |||||
1147 GAATGTGTGAGTTCGGAGCTTCTTTGAAACTTGGCATCTGGAGAGATGACCTTTTGAT 1206
QY |||||
1207 TCTGACTTTTCTTGTCCAGTCAAACCTTCATAGAGATCCCAAGTGTCTTCAAGATATCAC 1266
DB |||||
1207 TCTGACTTTTCTTGTCCAGTCAAACCTTCATAGAGATCCCAAGTGTCTTCAAGATATCAC 1266
QY |||||
1267 TGTACTCATGAAGCTGTGTATCATAGTGTGGGCTAACTTGGATTCATATAAATTACAAA 1326
DB |||||
1267 TGTACTCATGAAGCTGTGTATCATAGTGTGGGCTAACTTGGATTCATATAAATTACAAA 1326
QY |||||
1327 TTTCTTGGATCAGATGAAAGAGATAAGATAGTTTACAGGAATCTCTACAGAAACAGAAA 1386
DB |||||
1327 TTTCTTGGATCAGATGAAAGAGATAAGATAGTTTACAGGAATCTCTACAGAAACAGAAA 1386
QY |||||
1387 TGTCTTGTGTAACACATCCTTTGTACGAGGCCATTTGCCCTGCAGCAGCAGCTCCAATT 1446
DB |||||
1387 TGTCTTGTGTAACACATCCTTTGTACGAGGCCATTTGCCCTGCAGCAGCAGCTCCAATT 1446
QY |||||
1447 CGAGATTTTGGATTTGATCTGACATTTCTGGGACCCAGATGATCTGCATCCAGTACC 1506
DB |||||
1447 CGAGATTTTGGATTTGATCTGACATTTCTGGGACCCAGATGATCTGCATCCAGTACC 1506
QY |||||
1507 TATGAATGCTCATATGCGCGTTTATTAAGTACAGTCCATCCAGCGTCTCCTCCCTGCTT 1566
DB |||||
1507 TATGAATGCTCATATGCGCGTTTATTAAGTACAGTCCATCCAGCGTCTCCTCCCTGCTT 1566
QY |||||
1567 TGTACTCAGAGAGATGTTGAAAGTGGCAGAGTCTTCCCTCCGTTCTGGCTGAAACCCCA 1626
DB |||||
1567 TGTACTCAGAGAGATGTTGAAAGTGGCAGAGTCTTCCCTCCGTTCTGGCTGAAACCCCA 1626
QY |||||
1627 GATTCCTTTGAAAAGCATATTTAGAGAGATTTTGAACGTTGCTTCCCAATCCAGCATTT 1686
DB |||||
1627 GATTCCTTTGAAAAGCATATTTAGAGAGATTTTGAACGTTGCTTCCCAATCCAGCATTT 1686
QY |||||
1687 TTGAAGCTTCTGAAAAGCATATGCTCCTCCTGTAAGATGCTTCCAGTCTCCTCAGC 1746
DB |||||
1687 TTGAAGCTTCTGAAAAGCATATGCTCCTCCTGTAAGATGCTTCCAGTCTCCTCAGC 1746
QY |||||
1747 AGAGCCACCCAGGTGTTTCCAGAGAGAGTCAATTTCTCAACAGGACTTGGCAAGAGGAG 1806
DB |||||
1747 AGAGCCACCCAGGTGTTTCCAGAGAGAGTCAATTTCTCAACAGGACTTGGCAAGAGGAG 1806
QY |||||
1807 ATTACGCGAGGCTCAAAATTTATTTGACCCAAAAAAGAAACAATAGAGATCTCAGT 1866
DB |||||
1807 ATTACGCGAGGCTCAAAATTTATTTGACCCAAAAAAGAAACAATAGAGATCTCAGT 1866
QY |||||
1867 TATTCGAGAGAGAGAGAAAGTCTCGGGGAAATGCGGTGAGCGTTTACCTCACAATAT 1926
DB |||||
1867 TATTCGAGAGAGAGAGAAAGTCTCGGGGAAATGCGGTGAGCGTTTACCTCACAATAT 1926
QY |||||
1927 GAGGAGCTTAAAGAAAAAACAAGAGATATCATGAAACAGATGAAAAAACTACTTCCAGT 1986
DB |||||

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 20, 2005, 08:05:58 ; Search time 405 Seconds  
(without alignments)  
9607.570 Million cell updates/sec

Title: US-09-684-890A-1

Perfect score: 2378

Sequence: 1 gataaaccacagacacaa.....tataaaagtgtttgatg 2378

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.\*

1: /cgn2\_6/ptodata/1/ina/5A COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PCTUS COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2341.6	98.5	2392	4	US-09-620-312D-261
2	2312	97.2	2343	4	US-09-949-016-4105
3	315.4	13.3	37565	4	US-09-949-016-15847
4	174	7.3	601	4	US-09-949-016-145422
5	174	7.3	601	4	US-09-949-016-145423
6	174	7.3	601	4	US-09-949-016-145424
7	168	7.1	8374	4	US-09-949-016-11911
8	167	7.0	601	4	US-09-949-016-145428
9	138	5.8	601	4	US-09-949-016-145421
10	105.6	4.4	601	4	US-09-949-016-145431
11	88	3.7	601	4	US-09-949-016-145425
12	87.4	3.7	601	4	US-09-949-016-145426
13	87	3.7	601	4	US-09-949-016-145430
14	76.4	3.2	601	4	US-09-949-016-145432
15	61.2	2.6	7218	1	US-08-232-463-14
16	45	1.9	1702	4	US-09-799-451-574
17	40.2	1.7	832	4	US-09-621-976-2813
18	39.8	1.7	1158	4	US-09-710-279-69
19	39.8	1.7	1353	4	US-09-710-279-1919
20	39.8	1.7	1566	3	US-09-134-801C-2802
21	39.8	1.7	3055	4	US-09-710-279-3748
22	39.8	1.7	3226	4	US-09-710-279-4320
23	39.8	1.7	3229	4	US-09-710-279-4234
24	38.2	1.6	832	4	US-09-621-976-2813
25	38.2	1.6	1141	4	US-09-806-708B-22
26	38.2	1.6	1689	4	US-09-902-540-6020
27	38.2	1.6	1722	4	US-09-902-540-232

C	28	38	1.6	399	4	US-09-621-976-8976	Sequence 8976, Ap
	29	37.8	1.6	399	4	US-09-621-976-8976	Sequence 8976, Ap
	30	37.8	1.6	1141	4	US-09-806-708B-22	Sequence 22, Appl
	31	37.8	1.6	1916	4	US-09-976-594-900	Sequence 900, Appl
	32	36.6	1.5	4826	4	US-09-772-304A-1	Sequence 1, Appl
	33	36.6	1.5	256171	4	US-09-949-016-12822	Sequence 12822, A
	34	36.6	1.5	256176	4	US-09-949-016-15524	Sequence 15524, A
	35	36.4	1.5	140725	4	US-09-949-016-17074	Sequence 17074, A
C	36	36.4	1.5	138632	4	US-09-949-016-12781	Sequence 12781, A
	37	36.4	1.5	138637	4	US-09-949-016-17393	Sequence 17393, A
C	38	36.2	1.5	601	4	US-09-949-016-22174	Sequence 22174, A
	39	36.2	1.5	601	4	US-09-949-016-139679	Sequence 139679, A
	40	36.2	1.5	44653	4	US-09-949-016-11944	Sequence 11944, A
C	41	36.2	1.5	44653	4	US-09-949-016-15690	Sequence 15690, A
	42	36.2	1.5	1830121	4	US-09-557-884-1	Sequence 1, Appl
	43	36.2	1.5	1830121	4	US-09-643-990A-1	Sequence 1, Appl
	44	36	1.5	37822	4	US-09-949-016-16291	Sequence 16291, A
	45	35.8	1.5	945	4	US-09-902-540-8070	Sequence 8070, Ap

## ALIGNMENTS

RESULT 1  
US-09-620-312D-261  
; Sequence 261, Application US/09620312D  
; Patent No. 6569662  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Liu, Chenghua  
; APPLICANT: Asundi, Vinod  
; APPLICANT: Zhang, Jie  
; APPLICANT: Ren, Feiyan  
; APPLICANT: Chen, Rui-hong  
; APPLICANT: Zhao, Qing A.  
; APPLICANT: Wehrman, Tom  
; APPLICANT: Xue, Aidong J.  
; APPLICANT: Yang, Yonghong  
; APPLICANT: Wang, Jian-Rui  
; APPLICANT: Zhou, Ping  
; APPLICANT: Ma, Yungqing  
; APPLICANT: Wang, Dunrui  
; APPLICANT: Wang, Zhiwei  
; APPLICANT: John Tillinghast  
; APPLICANT: Drmanac, Radoje T.  
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and  
; FILE REFERENCE: 784CIP2B  
; CURRENT APPLICATION NUMBER: US/09/620,312D  
; PRIOR FILING DATE: 2000-07-19  
; PRIOR APPLICATION NUMBER: 09/552,317  
; PRIOR FILING DATE: 2000-04-25  
; PRIOR APPLICATION NUMBER: 09/488,725  
; PRIOR FILING DATE: 2000-01-21  
; NUMBER OF SEQ ID NOS: 1105  
; SOFTWARE: pt FL\_genes Version 1.0  
; SEQ ID NO 261  
; LENGTH: 2392  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (49)..(2274)  
US-09-620-312D-261

Query Match 98.5%; Score 2341.6; DB 4; Length 2392;  
Best Local Similarity 99.2%; Pred. No. 0;  
Matches 2353; Conservative 0; Mismatches 19; Indels 0; Gaps 0;  
Qy 7 CCCCAAGACACAAACATACCTTTTCGAGCAGTTGGCCCAAGATGGCGCCGCGGCGGA 66  
Db 7 CCCTCGAGGCCAAGATTTCGACAGCAGATGGCGCCGCGGCGGCGGCGGCGGGA 66

QY	67	CCGCTGGCGACGCGGAGCTGTGGCAGACCTGGCTTCTTAACACAGTCGTGTCTTTCGCG	126
Db	67	CCGCTGGCGACGCGGAGCTGTGGCAGACCTGGCTTCTTAACACAGTCGTGTCTTTCGCG	126
QY	127	CTCCGGGAGGACGTAAAAACGAGAGTCACACGAAAGCTGAGAAAACGAGCTTCTCGTCG	186
Db	127	CTCCGGGAGGACGTAAAAACGAGAGTCACACGAAAGCTGAGAAAACGAGCTTCTCGTCG	186
QY	187	TTGCTCTTCGTCGCGCGCGCGAGTTGCTGACGAGAAAACGCTGCTCTTTCGCGCGGA	246
Db	187	TTGCTCTTCGTCGCGCGCGCGAGTTGCTGACGAGAAAACGCTGCTCTTTCGCGCGGA	246
QY	247	GAGCTTTTTCTGTGGGACGAGAGACAGCTCTCTTCTTAGTCGTTCGCTTTCGCGGCCCC	306
Db	247	GAGCTTTTTCTGTGGGACGAGAGACAGCTCTCTTCTTAGTCGTTCGCTTTCGCGGCCCC	306
QY	307	AGCGGCGGCGGAGAGCGCGCGCTGTGCCAGTACAGAGATTGCTTTGCGATAAATCCA	366
Db	307	AGCGGCGGCGGAGAGCGCGCGCTGTGCCAGTACAGAGATTGCTTTGCGATAAATCCA	366
QY	367	CCCTCTGTTGAAATCTATCAAGTCTTGTAAAGCCCAACACATCATGTAGCACATTATA	426
Db	367	CCCTCTGTTGAAATCTATCAAGTCTTGTAAAGCCCAACACATCATGTAGCACATTATA	426
QY	427	GGAAATAAAAGGACTTATGGTATTAGAAATTACCTTAAAGATGGGGGAGAAATCTGAATTT	486
Db	427	GGAAATAAAAGGACTTATGGTATTAGAAATTACCTTAAAGATGGGGGAGAAATCTGAATTT	486
QY	487	GAAGGTGGAATAATCAACAGTGAATTTGTAGTACCACTCCAGTTCGGGAGAGATTTTTCACC	546
Db	487	GAAGGTGGAATAATCAACAGTGAATTTGTAGTACCACTCCAGTTCGGGAGAGATTTTTCACC	546
QY	547	AGTTCCACCTCTCTGACTCTTAAAGCATGCTGCAATGATCCCAAGTGAATCTCTGGATCCC	606
Db	547	AGTTCCACCTCTCTGACTCTTAAAGCATGCTGCAATGATCCCAAGTGAATCTCTGGATCCC	606
QY	607	CAGTGTGCTGTAAACATCAGACAAACGTAATCAGAAATTTACTCTCTAGTGGCGCGAG	666
Db	607	CAGTGTGCTGTAAACATCAGACAAACGTAATCAGAAATTTACTCACTAGTGGCGCGAG	666
QY	667	ACACCACTAAACGTGATATATCTTTAGAAAGCCGAAAGAGAAAGTCTAGTACTCAATAAA	726
Db	667	ACACCACTAAACGTGATATATCTTTAGAAAGCCGAAAGAGAAAGTCTAGTACTCAATAAA	726
QY	727	GGAAAGGCGTATACCCGATCTCTAGGAGAGACAGTGTGCAATTCACATTTGGGCGCATTG	786
Db	727	GGAAAGGCGTATACCCGATCTCTAGGAGAGACAGTGTGCAATTTGACATTTGGGCGCATTG	786
QY	787	GACGCACTCCCAAGACTCTATTGGAACAAAACGCAAGATGAAGTGTGGCATACCCA	846
Db	787	GCAGCAGTCCCAAGACTCTATTGGAACAAAACGCAAGATGAAGTGTGGCATACCCA	846
QY	847	CTGTACATCTTATGAAATGAGAGACTTTTCTGACATACATCATAGTCTGTGTACACAGC	906
Db	847	CTGTACATCTTATGAAATGAGAGACTTTTCTGACATACATCATAGTCTGTGTACACAGC	906
QY	907	CCTGGAATAATTTGGAAGCTGTGGTCCATTTGCCCATGCACTCGCGCTGAAAGATAAC	966
Db	907	CCTGGAATAATTTGGAAGCTGTGGTCCATTTGCCCATGCACTCGCGCTGAAAGATAAC	966
QY	967	TATGTTATGATCGTGTGCTGACTCTGTCTACCTGTGTCCCAATATCTTAGTGATC	1026
Db	967	TATGTTATGATCGTGTGCTGACTCTGTCTACCTGTGTCCCAATATCTTAGTGATC	1026
QY	1027	GCTACTGAATCAGGAATGCTGTATCATCTGTGCTGTAGAGGGGAGAGAAAGATGAC	1086
Db	1027	GCTACTGAATCAGGAATGCTGTATCATCTGTGCTGTAGAGGGGAGAGAAAGATGAC	1086
QY	1087	CACAGCTCAGAAAGTCTGGGATTCAGGATTGACCTCATTCCTCTCTGTATGTGTTT	1146
Db	1087	CACAGCTCAGAAAGTCTGGGATTCAGGATTGACCTCATTCCTCTCTGTATGTGTTT	1146
QY	1147	GAAATGTTGAGTTGGAGCTTGCTTTTGAACCTGGCATCTGGAGGATGACCCCTTTTGAT	1206

Db	1147	GAAATGTTGAGTTGGAGCTTGCTTTGAAACCTGGCATCTGGAGGATGACCCCTTTTGAT	1206
QY	1207	TCGTGACTTTCTGTGCCAGTCAAACTTCATAGAGATCCCAAGTGTCTTCAAGATATCAC	1266
Db	1207	TCGTGACTTTCTGTGCCAGTCAAACTTCATAGAGATCCCAAGTGTCTTCAAGATATCAC	1266
QY	1267	TGTACTCATGAAGCTGGTGTACATAGTGTGGGCTAACTTTGGATTATATAAACTTCAAAA	1326
Db	1267	TGTACTCATGAAGCTGGTGTACATAGTGTGGGCTAACTTTGGATTATATAAACTTCAAAA	1326
QY	1327	TTTCTTTGGATCAGATGAAGAAGATAGGATAGTTTACAGAACTCTCTACAGAACGAAA	1386
Db	1327	TTTCTTTGGATCAGATGAAGAAGATAGGATAGTTTACAGAACTCTCTACAGAACGAAA	1386
QY	1387	TGCTTTGTTGAACACATCTTTGTACGAGGCCATTTGCCCTGCAGGCGAGCTCCAATTT	1446
Db	1387	TGCTTTGTTGAACACATCTTTGTACGAGGCCATTTGCCCTGCAGGCGAGCTCCAATTT	1446
QY	1447	CGAGGATTTTGGATTGTAGCTGACATTTCTGGGACCCACGATGATCTGCACTCACAGTACC	1506
Db	1447	CGAGGATTTTGGATTGTAGCTGACATTTCTGGGACCCACGATGATCTGCACTCACAGTACC	1506
QY	1507	TATGAATGCTCATATGCGCGTTATTAACTACAGTCCATCCAGCGTCTCTCCCTGCTT	1566
Db	1507	TATGAATGCTCATATGCGCGTTATTAACTACAGTCCATCCAGCGTCTCTCCCTGCTT	1566
QY	1567	TGTACTCGAGAGATGTTGAAGTGGCAGAGTCTTCCCTCGGTCTTGGCTGAAACCCCA	1626
Db	1567	TGTACTCGAGAGATGTTGAAGTGGCAGAGTCTTCCCTCGGTCTTGGCTGAAACCCCA	1626
QY	1627	GATTCCTTTGAAAAGCATATTAGAACATTTTGCACAGTAGTGTGCCAATCCAGCATTT	1686
Db	1627	GATTCCTTTGAAAAGCATATTAGAACATTTTGCACAGTAGTGTGCCAATCCAGCATTT	1686
QY	1687	TTGAAAGCTTCTGAAAAGGACATAGCCCTCTCTCTGAAAGATGCTTCCAGTCTCTCAGC	1746
Db	1687	TTGAAAGCTTCTGAAAAGGACATAGCCCTCTCTCTGAAAGATGCTTCCAGTCTCTCAGC	1746
QY	1747	AGAGCCACCCAGGTGTTACAGAGCAGTACATTTCTCAACAGGACTTTGGCAAGGAGGAG	1806
Db	1747	AGAGCCACCCAGGTGTTACAGAGCAGTACATTTCTCAACAGGACTTTGGCAAGGAGGAG	1806
QY	1807	ATTACGCGAGGCTCAAAATTTATTGTGACCAAAAAAGAAAACAACTAGAAGATCTCAGT	1866
Db	1807	ATTACGCGAGGCTCAAAATTTATTGTGACCAAAAAAGAAAACAACTAGAAGATCTCAGT	1866
QY	1867	TATTTGCGAGAGAGAGGAAAAAGTCTGCGGGAATGGCTGAGCGTTTAGCTGACAAATAT	1926
Db	1867	TATTTGCGAGAGAGAGGAAAAAGTCTGCGGGAATGGCTGAGCGTTTAGCTGACAAATAT	1926
QY	1927	GAGGAGCTTAAAGNAAAACAAGGATATCATGACAGGATGAAAACCTACTTCAAGT	1986
Db	1927	GAGGAGCTTAAAGNAAAACAAGGATATCATGACAGGATGAAAACCTACTTCAAGT	1986
QY	1987	TTTCACTCTGAGCTCCAGTCTCTCTGATAGTACGAGACATGAAGAAAAGATTTACAG	2046
Db	1987	TTTCACTCTGAGCTCCAGTCTCTCTGATAGTACGAGACATGAAGAAAAGATTTACAG	2046
QY	2047	CTGATACCTGATCAACTTCGACATTTGGGCAATGCCATCAACAGGTTTACTATGAAAAAG	2106
Db	2047	CTGATACCTGATCAACTTCGACATTTGGGCAATGCCATCAACAGGTTTACTATGAAAAAG	2106
QY	2107	GATTATCAACGAAAGATGGAGAGGTTGAGTCTTCCAAAACCCACCATTTATTTCTC	2166
Db	2107	GATTATCAACGAAAGATGGAGAGGTTGAGTCTTCCAAAACCCACCATTTATTTCTC	2166
QY	2167	AGTGCCCTACAGCAAAAGTGCATTCAGTCCATCTGAAAAGAGAGGGTGAACATATAAGG	2226
Db	2167	AGTGCCCTACAGCAAAAGTGCATTCAGTCCATCTGAAAAGAGAGGGTGAACATATAAGG	2226
QY	2227	GAAATGTTGAGCAAAATCAATGATATTCGCAATCATGTAAACTTTCTGACACCAACGAGG	2286
Db	2227	GAAATGTTGAGCAAAATCAATGATATTCGCAATCATGTAAACTTTCTGACACCAACGAGG	2286

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 20, 2005, 04:09:16 ; Search time 10297 Seconds  
(without alignments)  
11190.298 Million cell updates/sec

Title: US-09-684-890A-1  
Perfect score: 2378  
Sequence: 1 gataaccacacagacacaa.....tataaaagggtgtttgatg 2378

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl.\*

1: gb\_ba.\*  
2: gb\_hgt.\*  
3: gb\_in.\*  
4: gb\_om.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_ey.\*  
13: gb\_un.\*  
14: gb\_vi.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2378	100.0	2378	9 HSNUP88	Y08612 Homo sapien
2	2341.6	98.5	2392	6 AR338770	AR338770 Sequence
3	2315.8	97.4	2360	6 CQ725479	CQ725479 Sequence
4	2301	96.8	2357	9 BC000335	BC000335 Homo sapi
5	2301	96.8	2366	6 AX828393	AX828393 Sequence
6	2287.8	95.4	2386	11 BV179863	BV179863 sqmml0705
7	1721	72.4	2418	10 BC032929	BC032929 Mus muscu
8	1703.8	71.6	2412	10 RN093692	U93692 Rattus norv
9	1700.4	71.5	2433	10 BC072524	BC072524 Rattus no
10	1698.8	71.4	2310	10 MMU532593	AJ532593 Mus muscu
11	945.4	39.8	2393	5 AJ617672	AJ617672 Xenopus 1
12	905.8	38.1	2735	5 BC077397	BC077397 Xenopus 1
13	904.2	38.0	2478	5 AJ617673	AJ617673 Mus muscu
C 14	630.2	26.5	197577	2 AC124124	AC124124 Mus muscu
C 15	630.2	26.5	228883	10 AC127173	AC127173 Mus muscu
C 16	597.4	25.1	601	11 BV168085	BV168085 sqmml7240
17	523.8	22.0	746	4 AY553927	AY553927 Sus scrof
18	408	17.2	651	10 U01135	U01135 Mus musculu
C 19	315.4	13.3	118276	9 AC004148	AC004148 Homo sapi

C 20	217	9.1	250275	2	AC095695	AC095695 Rattus no
C 21	213.8	9.0	180274	2	AC027185	AC027185 Mus muscu
C 22	212.2	8.9	254116	10	AL596136	AL596136 Mouse DNA
C 23	200.6	8.4	201	11	BV198093	BV198093 sqmml9650
C 24	200.6	8.4	201	11	BV200393	BV200393 sqmml20405
C 25	200.6	8.4	201	11	BV200394	BV200394 sqmml20405
C 26	170	7.1	2344	3	AK112449	AK112449 Ciona int
C 27	167.6	7.0	560	11	BV178251	BV178251 sqmml99469
C 28	166	7.0	466	6	CQ070585	CQ070585 Sequence
C 29	166	7.0	466	6	CQ099041	CQ099041 Sequence
C 30	166	7.0	466	6	CQ137971	CQ137971 Sequence
C 31	166	7.0	466	6	CQ175148	CQ175148 Sequence
C 32	166	7.0	466	6	CQ221345	CQ221345 Sequence
C 33	166	7.0	466	6	CQ259446	CQ259446 Sequence
C 34	166	7.0	466	6	CQ297165	CQ297165 Sequence
C 35	166	7.0	466	6	CQ333529	CQ333529 Sequence
C 36	165.4	7.0	187	6	CQ079765	CQ079765 Sequence
C 37	165.4	7.0	187	6	CQ112093	CQ112093 Sequence
C 38	165.4	7.0	187	6	CQ150910	CQ150910 Sequence
C 39	165.4	7.0	187	6	CQ185006	CQ185006 Sequence
C 40	165.4	7.0	187	6	CQ234252	CQ234252 Sequence
C 41	165.4	7.0	187	6	CQ271955	CQ271955 Sequence
C 42	165.4	7.0	187	6	CQ309582	CQ309582 Sequence
C 43	165.4	7.0	187	6	CQ346195	CQ346195 Sequence
C 44	153	6.4	455	11	G27792	G27792 human STS S
C 45	149	6.3	546	6	AX780812	AX780812 Sequence

#### ALIGNMENTS

RESULT 1  
HSNUP88  
LOCUS Homo sapiens mRNA for nuclear pore complex protein 88 (Nup88 gene).  
DEFINITION Homo sapiens mRNA for nuclear pore complex protein 88 (Nup88 gene).  
ACCESSION Y08612  
VERSION Y08612.2 GI:5541878  
KEYWORDS nuclear pore complex protein; Nup88 gene.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1  
AUTHORS Fornerod, M., van Deursen, J., van Baal, S., Reynolds, A., Davis, D., Murti, K.G., Franssen, J. and Grosveld, G.  
TITLE The human homologue of yeast CRM1 is in a dynamic subcomplex with CAN/Nup214 and a novel nuclear pore component Nup88  
JOURNAL EMBO J. 16 (4), 807-816 (1997)  
MEDLINE 97201523  
REFERENCE 2  
AUTHORS Fornerod, M.  
TITLE Direct Submission  
JOURNAL Submitted (02-OCT-1996) M. Fornerod, St. Jude Children's Research Hospital, Department of Genetics, 332 N. Lauderdale, Memphis, TN 38105, USA  
REMARK Revised by [3]  
REFERENCE 3 (bases 1 to 2378)  
AUTHORS Fornerod, M.  
TITLE Direct Submission  
JOURNAL Submitted (19-JUL-1999) M. Fornerod, St. Jude Children's Research Hospital, Department of Genetics, 332 N. Lauderdale, Memphis, TN 38105, USA  
COMMENT On Jul 20, 1999 this sequence version replaced gi:1707521.  
FEATURES  
source Location/Qualifiers  
1. .2378  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/chromosome="17"  
/map="17p13"  
/tissue\_type="placenta"  
/clone\_lib="Clontech Hu2002B#29203"



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 20, 2005, 03:21:30 ; Search time 1279 Seconds  
(without alignments)  
11006.368 Million cell updates/sec

Title: US-09-684-890A-1  
Perfect score: 2378  
Sequence: 1 gataaacccacaaagacacaa.....tataaaagggtgtttgatg 2378

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N Geneseq\_16Dec04:.\*  
1: geneseqn1980e:.\*  
2: geneseqn1990e:.\*  
3: geneseqn2000e:.\*  
4: geneseqn2001as:.\*  
5: geneseqn2001bs:.\*  
6: geneseqn2002as:.\*  
7: geneseqn2002bs:.\*  
8: geneseqn2003as:.\*  
9: geneseqn2003bs:.\*  
10: geneseqn2003cs:.\*  
11: geneseqn2003ds:.\*  
12: geneseqn2004as:.\*  
13: geneseqn2004bs:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2347	98.7	2464	AAK52959	AAK52959 Human pol
2	2343.8	98.6	2464	AAK52960	AAK52960 Human pol
3	2341.6	98.5	2392	AAI58382	AAI58382 Human pol
4	2341.6	98.5	2392	ADQ98591	ADQ98591 DNA encod
5	2341.6	98.5	2392	ADQ98591	ADQ98591 Novel hum
6	2336	98.2	2465	AAI60168	AAI60168 Human pol
7	2330.6	98.0	2393	AAK51976	AAK51976 Human pol
8	2330.6	98.0	2393	AAK51975	AAK51975 Human pol
9	2315	97.4	2346	ACN40104	ACN40104 Tumour-as
10	2301	96.8	2366	ADK60464	ADK60464 Angiogene
11	2301	96.8	2366	ADK60765	ADK60765 Angiogene
12	2301	96.8	2366	ADP73087	ADP73087 Angiogene
13	1710.2	71.9	2229	ADJ67702	ADJ67702 Kflic DNA
14	1703.8	71.6	2412	ADB59152	ADB59152 Toxicity
15	1703.8	71.6	2412	ADP72905	ADP72905 Renal tox
16	393.4	16.5	456	ACH28622	ACH28622 Human adu
17	322.2	13.5	429	ACH21235	ACH21235 Human adu
18	315.4	13.3	1084	ACN40103	ACN40103 Tumour-as
19	315.4	13.3	65854	AAK86282	AAK86282 Human imm
20	243.4	10.2	514	ACH69405	ACH69405 Human gen

C 21	168	7.1	32207	5	ABA19666	ABA19666 Human ncr
C 22	166	7.0	466	4	AAI16452	AAI16452 Probe #83
C 23	166	7.0	466	4	ABA59402	ABA59402 Human foe
C 24	166	7.0	466	4	AAI39214	AAI39214 Probe #79
C 25	166	7.0	466	4	ABA28078	ABA28078 Probe #65
C 26	166	7.0	466	4	AAK33436	AAK33436 Human bon
C 27	166	7.0	466	4	AAK07632	AAK07632 Human bra
C 28	166	7.0	466	4	ABA33194	ABA33194 Human liv
C 29	166	7.0	466	6	ABA08279	ABA08279 Human gen
C 30	165.4	7.0	187	4	AAI25632	AAI25632 Probe #15
C 31	165.4	7.0	187	4	ABA71911	ABA71911 Human foe
C 32	165.4	7.0	187	4	AAI52266	AAI52266 Probe #20
C 33	165.4	7.0	187	4	ABA37936	ABA37936 Probe #16
C 34	165.4	7.0	187	4	AAK46375	AAK46375 Human bon
C 35	165.4	7.0	187	4	AAK20298	AAK20298 Human bra
C 36	165.4	7.0	187	4	ABA46101	ABA46101 Human liv
C 37	165.4	7.0	187	6	ABA20696	ABA20696 Human gen
C 38	149	6.3	546	10	ADF82413	ADF82413 Leukaemia
C 39	146.4	6.2	212	3	ADF57323	ADF57323 Urogenita
C 40	145	6.1	147	12	ACH83105	ACH83105 Human gen
C 41	136.8	5.8	583	13	ADQ54230	ADQ54230 Novel can
C 42	134.6	5.7	459	4	AAI16224	AAI16224 Probe #61
C 43	134.6	5.7	459	4	ABA58929	ABA58929 Human foe
C 44	134.6	5.7	459	4	AAI38646	AAI38646 Probe #73
C 45	134.6	5.7	459	4	ABA27796	ABA27796 Probe #82

## ALIGNMENTS

## RESULT 1

AAK52959  
ID AAK52959 standard; cDNA; 2464 BP.  
XX  
AC AAK52959;  
XX  
DT 06-NOV-2001 (first entry)  
XX  
DE Human polynucleotide SEQ ID NO 2488.

Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
tissue growth factor; immunomodulatory; cancer; leukaemia;  
nervous system disorder; arthritis; inflammation; ss.

OS Homo sapiens.

PN WO200157190-A2.

XX

PD 09-AUG-2001.

XX

PF 05-FEB-2001; 2001WO-US004098.

XX

PR 03-FEB-2000; 2000US-00496914.

PR 27-APR-2000; 2000US-00560875.

PR 20-JUN-2000; 2000US-00598075.

PR 19-JUL-2000; 2000US-00620325.

PR 01-SEP-2000; 2000US-00654936.

PR 15-SEP-2000; 2000US-00663561.

PR 20-OCT-2000; 2000US-00693325.

PR 30-NOV-2000; 2000US-00728422.

XX (HYSE-) HYSEQ INC.

PA

XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;

PI Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;

PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;

XX WPI; 2001-476283/51.

DR P-PSDB; AAM79826.

XX

PT Nucleic acids encoding polypeptides with cytokine-like activities, useful

PT in diagnosis and gene therapy.

